

# Current Topics in Nonstationary Analysis

Proceedings of the Second Workshop on Nonstationary  
Random Processes and their Applications

San Diego, CA    June 11 – 12, 1995

DTIC QUALITY INSPECTED 4

*edited by*

**George Treviño**

*CHIRES Associates, Inc.*

**Jay Hardin**

*Nasa-Langley Research Center*

**Bruce Douglas**

*Naval Surface Warfare Center-Carderock*

**Edgar Andreas**

*US Army Cold Regions Research and Engineering Laboratory*

This document has been approved  
for public release and sale; its  
distribution is unlimited.



**World Scientific**

*Singapore • New Jersey • London • Hong Kong*

19961230 026

*Published by*

World Scientific Publishing Co. Pte. Ltd.

P O Box 128, Farrer Road, Singapore 912805

USA office: Suite 1B, 1060 Main Street, River Edge, NJ 07661

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

**British Library Cataloguing-in-Publication Data**

A catalogue record for this book is available from the British Library.

**CURRENT TOPICS IN NONSTATIONARY ANALYSIS**

**Proceedings of the Second Workshop on Nonstationary Random Processes  
and their Applications**

Copyright © 1996 by World Scientific Publishing Co. Pte. Ltd.

*All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.*

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA.

ISBN 981-02-2703-5

Printed in Singapore.

REPORT DOCUMENTATION PAGE			Form Approved OMB NO. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comment regarding this burden estimates or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.				
1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE December 1996	3. REPORT TYPE AND DATES COVERED Final 1 June 1995 - 31 May 1996		
4. TITLE AND SUBTITLE Current Topics in Nonstationary Analysis		5. FUNDING NUMBERS DAAH04-95-1-0406		
6. AUTHOR(S) George Trevino (principal investigator)		8. PERFORMING ORGANIZATION REPORT NUMBER		
7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(ES) Chires, Inc. Houghton, MI 49931				
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Office P.O. Box 12211 Research Triangle Park,, NC 27709-2211		10. SPONSORING / MONITORING AGENCY REPORT NUMBER ARO 32835.1-GS-CF		
11. SUPPLEMENTARY NOTES The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution unlimited.		12 b. DISTRIBUTION CODE		
13. ABSTRACT (Maximum 200 words)  ABSTRACT NOT FURNISHED				
14. SUBJECT TERMS		15. NUMBER IF PAGES		
		16. PRICE CODE		
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UL	

# **Current Topics in Nonstationary Analysis**

Proceedings of the Second Workshop on Nonstationary  
Random Processes and their Applications

## PREFACE

Nonstationarity is another name for intermittency, a phenomenon which affects many physical processes in the atmospheric boundary layer. Among these are the transfers of heat, momentum, and moisture, and thus the propagation of electromagnetic waves. Nonstationarity also appears in the nonlinear propagation of acoustic waves and in noise radiation by supersonic jets and helicopters. Such signals are commonplace in military communication systems, ship and submarine stealth, and maneuvering and control problems. Also, data collected in many diverse R&D programs frequently exhibit nonstationary features. Of these a common one is turbulence data, which we know from observations to be a complicated nonlinear dynamical phenomenon of limited predictability.

Several advances in the theory of nonstationary random processes have been made since the first workshop convened in 1991. And although much research remains yet to be done, it seems opportune to convert some of the existing work into a more permanent form. It is debatable what impact the new knowledge has had on the development of practical models of this important non-equilibrium phenomenon. The chief aim of the workshop was therefore to provide a forum at which the recent important contributions could be reported and discussed among researchers from government, academia, and industry.

The theme of the workshop was all aspects of nonstationary analysis, with the appeal for participation being made to engineers, scientists, and mathematicians alike. This appeal is consistent with that of the first workshop. The intent was to create a diverse environment for researchers working in this genuinely multidisciplinary field to mutually share their ideas. The premier objective of the workshop was to consolidate recent developments in nonstationary analysis and present the material in a "tutorial mode." A second objective was to delineate open problems.

We would like to express our gratitude to those who assisted us in convening the workshop. Sincere thanks go to Jack Preisser (NASA Langley), Frank Halsall (NSWC-Carderock), and Walter Bach (Army Research Office), who collectively provided most of the financial support necessary to bring the workshop to fruition. We gratefully acknowledge the administrative support provided by Norma Treviño. To the anonymous referees who reviewed the submitted manuscripts, and to those who served as session chairmen (Frank Halsall, Alan Piersol and Ken Bolland), we are also grateful. Most of all, though, we express our deep appreciation to the authors, whose hard work and dedication made the workshop a success.

George Treviño  
Jay Hardin  
Bruce Douglas  
Edgar Andreas

## CONTENTS

Preface	v
Matched Filtering in the Case of Nonstationary Interference <i>J. A. Sills and E. W. Kamen</i>	1
Inconsistencies in Parameters Estimated from Impulsive Noise <i>R. D. Pierce</i>	15
Detrending Turbulence Time Series with Wavelets <i>E. L. Andreas and G. Treviño</i>	35
A Two Dimensional Power Spectral Study for Some Nonstationary Processes <i>G. L. Smith and A. G. Miamee</i>	75
Multifractal Characterizations of Intermittency in Nonstationary Geophysical Signals and Fields <i>A. B. Davis, A. L. Marshak, W. J. Wiscombe, and R. F. Cahalan</i>	97
Short-Time Averaging of Turbulent Shear Flow Signals <i>A. K. Gupta</i>	159
Scale-Invariant Formulation of Nonstationary Signals <i>G. Treviño</i>	175

## **MATCHED FILTERING IN THE CASE OF NONSTATIONARY INTERFERENCE**

J.A. Sills  
Department of Radiolocation Science  
Southwest Research Institute  
San Antonio, Texas 78228-0510  
E-mail: jsills@swri.edu

E.W. Kamen  
School of Electrical and Computer Engineering  
Georgia Institute of Technology  
Atlanta, Georgia 30332-0250  
E-mail: ekamen@ee.gatech.edu

### **ABSTRACT**

This paper deals with a large class of nonstationary stochastic processes generated by passing white noise through a general linear time-varying system. It is shown that such processes can be characterized in terms of a family of wide-sense stationary processes. This formulation is used to define the power spectrum for the nonstationary processes considered in the paper. Then the power spectrum is used to give a suboptimal solution to the matched filtering problem for the case of nonstationary interference. This suboptimal solution is shown to be nearly optimal under conditions corresponding to a sufficiently small rate of variation. A numerical example compares this suboptimal solution to an optimal solution.

### **1. Introduction**

Matched filters are used extensively in radar and communication systems to detect the presence of a signal buried in noise. For example in a radar application, a sinusoidal pulse is transmitted, reflected back from a target, and received. A matched filter is used to detect the arrival of the received signal, and the arrival time can be used to determine the target's distance. Early work on matched filters, dating back to the 1940's, treated the case of stationary interference and ideal (time-invariant) propagation channels<sup>1</sup>. Matched filtering in the case of nonideal (time-varying) channels (RAKE filtering) has received considerable attention since it was first treated by Price and Green<sup>2</sup> in 1958.

In recent years applications for detecting signals buried in nonstationary noise have been identified. For example, radar and communication systems affected by counter

measures must combat nonstationary interference. Recently Modestino and Melendez<sup>3</sup> presented a matched filter which treats nonstationary interference using techniques based a local covariance matrix. In this paper we construct a matched filter for the case of nonstationary interference using frequency-domain techniques.

We consider interference signals from the general class of nonstationary stochastic processes generated by passing white noise through a causal linear system as shown in Figure 1. More precisely, we consider the nonstationary process  $v(n)$  given by the superposition summation

$$v(n) = \sum_{m=-\infty}^n g(n, m) \varepsilon(m) \quad (1)$$

where  $\varepsilon(m)$  is zero-mean unit-variance white noise and  $g(n, m)$  is the linear system's impulse-response function defined as the system response at index  $n$  to an impulse applied at index  $m$ .

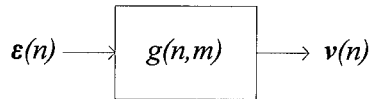


Figure 1. Block diagram of signal model.

We restrict our attention to processes with finite-valued autocorrelation functions  $r_v(i+n, i) = E[v(i+n)v(i)]$  by requiring that the impulse-response function  $g(n, m)$  be square summable over  $m$  for each value of  $n$ ; that is, there exists a function  $c(n)$ , with  $c(n) < \infty$ , such that

$$\sum_{m=-\infty}^n [g(n, m)]^2 < c(n) \quad (2)$$

We now formally define the class of processes considered in this paper.

**Definition 1.** A random process  $v(n)$  generated by (1), where the impulse-response function  $g(n, m)$  satisfies (2), is said to be purely nondeterministic (PND).

Cramer<sup>4</sup> extended the results of Wold by showing that any process with finite-valued mean and autocorrelation functions could be decomposed into its deterministic (as defined in [Cramer 1961]) and PND parts. Therefore, we are considering all such processes for which the deterministic component is zero.

The problem of detecting signals buried in nonstationary PND noise is described in Section 2. In Section 3 we show that associated with any PND nonstationary process is an associated family of jointly wide-sense stationary processes. This formulation leads to the notion of power spectrum for a class of PND processes, and it is also used to formulate a "clairvoyant" detection problem in Section 4. The optimal solution to the clairvoyant detection problem is given in terms of the power spectrum. In Section 5 the "clairvoyant" solution is applied to the detection problem of Section 2. This suboptimal



solution is shown to be nearly optimal under slowly varying conditions. A computer simulation presented in Section 6 compares the suboptimal solution with the optimal clairvoyant solution.

## 2. The Detection Problem

In the detection problem, a measured process  $r(n, n_d) = s(n - n_d) + v(n)$  is composed of a known signal  $s(n)$  and a corrupting interference  $v(n)$  as shown in Figure 2. The signal component  $s(n)$  is delayed by the unknown quantity  $n_d$ , and the interference  $v(n)$  is a nonstationary PND process with signal model given by the impulse-response function  $g(n, m)$ . The objective is to detect the signal  $s(n)$  and determine the most likely value for the delay  $n_d$  from the measured process  $r(n, n_d)$ .

The received signal  $r(n, n_d)$  is passed through a linear filter with impulse-response function  $h(n, m)$  to generate the response  $y(n, n_d)$  given by

$$y(n, n_d) = \sum_{m=-\infty}^n h(n, m) r(m, n_d) = y_s(n, n_d) + y_v(n)$$

where the noise component  $y_v(n)$  is given by

$$y_v(n) = \sum_{m=-\infty}^n h(n, m) v(m)$$

and the signal component  $y_s(n, n_d)$  is given by

$$y_s(n, n_d) = \sum_{m=-\infty}^n h(n, m) s(m - n_d)$$

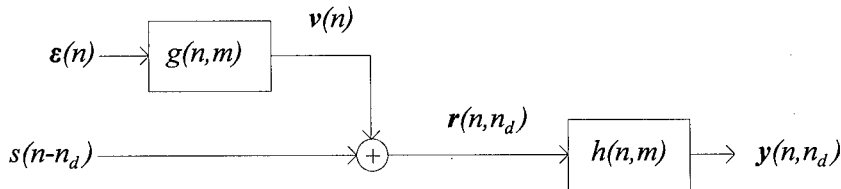


Figure 2. Block diagram of detection problem.

As in the stationary case, the detection filter is to compress the signal energy in time thereby generating a response that reaches a maximum amplitude at a particular time index  $n_p$ . The time index at which the maximum occurs is then used to find the delay  $n_d$ . Solving this problem requires finding the filter  $h_{MF}(n, m)$  such that with  $h(n, m) = h_{MF}(n, m)$  the signal-to-noise ratio defined by

$$R(n, n_d) = \frac{|y_s(n, n_d)|^2}{r_{y_v}(n, n)}$$

is maximized at the specific index  $n=n_p=n_d+n_0$  where  $n_0$  is a design parameter and

$$r_{y_v}(n, n) = E[y_v(n) y_v(n)]$$

The delay is then calculated as  $n_d=n_p-n_0$ .

It is well known that if the impulse-response function  $g(n, m)$  of the signal model for  $v(n)$  is time invariant so that  $v(n)$  is stationary, then the detection filter which maximizes the signal-to-noise ratio is given by

$$H_{MF}(e^{j\omega}) = \frac{S^*(e^{j\omega}) e^{-j\omega n_0}}{S_v(e^{j\omega})}$$

where  $H_{MF}(e^{j\omega})$  is the frequency-response function of the filter  $h_{MF}(n, m)$ ,  $S(e^{j\omega})$  is the Fourier transform of the known signal  $s(n)$ ,  $S_v(e^{j\omega})$  is the power spectrum of the stationary noise  $v(n)$ , and  $n_0$  is a design parameter chosen to insure a causal solution. Given the relationship between  $H_{MF}(e^{j\omega})$  and  $S(e^{j\omega})$  for the stationary case, the optimal detection filter is commonly called a matched filter.

Unfortunately, in many cases finding the optimal solution is extremely difficult. In this paper we seek a "viable" suboptimal solution. The suboptimal solution is based on the power spectrum defined the next section.

### 3. Power Spectra for PND Processes

In this section we define the power spectrum for a class of PND nonstationary processes. We begin by showing that every nonstationary PND process has associated with it a family of jointly wide-sense stationary processes.

Consider a PND nonstationary process  $v(n)$  with signal model given by the impulse-response function  $g(n, m)$ . Introduce the integer  $k$  by defining  $g_k(n) = g(k, k-n)$ . For a fixed integer  $k$ , let  $g_k(n)$  be the impulse-response function of a linear time-invariant (LTI) system. The collection of impulse-response functions  $\{g_k(n)\}$  forms a family of LTI systems indexed by  $k$ . A family of processes  $\{v_k(n)\}$  associated with the nonstationary process  $v(n)$  is defined by Sills<sup>5</sup> in terms of the  $\{g_k(n)\}$  by

$$v_k(n) = \sum_{i=-\infty}^n g_k(n-i) \varepsilon(i)$$

We can make three immediate observations regarding the associated family of processes. First, the nonstationary process  $v(n)$  can be calculated from the  $\{v_k(n)\}$  by evaluating the  $\{v_k(n)\}$  at  $k=n$ ; that is

$$v(n) = v_k(n) |_{k=n}$$

Second, the family of processes  $\{v_k(n)\}$  are jointly wide-sense stationary (WSS), and

hence, the family of crosscorrelation functions  $\{r_{v_k v_m}(n)\}$  is defined as

$$r_{v_k v_m}(n) = E[v_k(i+n) v_m(i)] \quad (3)$$

And finally third, the autocorrelation of the nonstationary process  $v(n)$  is related to the family of crosscorrelations  $\{r_{v_k v_m}(n)\}$  by

$$r_v(n, m) = r_{v_n v_m}(n-m)$$

The finiteness of  $r_{v_k v_m}(i+n, i)$  is insured by (2).

For the remainder of the paper we will consider PND processes for which the impulse-response function  $g(n, m)$  is absolutely summable in  $m$  for each  $n$ ; that is, we now assume that there exists a function  $c(n)$ , with  $c(n) < \infty$  for all  $n$ , such that

$$\sum_{m=-\infty}^n |g(n, m)| < c(n) \quad (4)$$

Consequently, we are now considering a subclass of PND processes since (4) is more constraining than (2). The additional constraint given in (4) insures the existence of the power spectrum defined below.

For any fixed integers  $k$  and  $m$ , the joint power spectrum between  $v_k(n)$  and  $v_m(n)$  is defined by

$$S_{v_k v_m}(e^{j\omega}) = \sum_{i=-\infty}^{\infty} r_{v_k v_m}(i) e^{-j\omega i} \quad (5)$$

where the crosscorrelation  $r_{v_k v_m}(n)$  is given by (3). The convergence of the sum in (5) is guaranteed by (4). We can now state the following result which is used in Section 5.

**Proposition 2.** The joint power spectrum  $S_{v_k v_m}(e^{j\omega})$  defined by (5) can be calculated using the relationship

$$S_{v_k v_m}(e^{j\omega}) = G(e^{j\omega}, k) G(e^{-j\omega}, m)$$

for any integers  $k$  and  $m$  where  $G(e^{j\omega}, k)$  is defined by

$$G(e^{j\omega}, k) = \sum_{i=0}^{\infty} g(k, k-i) e^{-j\omega i} \quad (6)$$

The quantity  $G(e^{j\omega}, k)$  defined by (6) is the frequency-response function of the signal model and was first studied by Zadeh<sup>6</sup> in 1950. Like the impulse-response function, it completely characterizes the signal model.

It follows from (5) that the autocorrelation function  $r_v(n, m)$  of the nonstationary process  $v(n)$  can be expressed as

$$r_v(n, m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_{v_n v_m}(e^{j\omega}) e^{j\omega(n-m)} d\omega \quad (7)$$

Equation (7) provides some justification for the following definition of the power spectrum for the nonstationary processes considered in this paper.

**Definition 3.** Given a PND nonstationary process  $v(n)$  with signal model  $g(n, m)$

satisfying (4), the power spectrum  $S_v(e^{j\omega}, n)$  of the nonstationary process  $v(n)$  is defined by

$$S_v(e^{j\omega}, n) = S_{v_n v_n}(e^{j\omega}) \quad (8)$$

where  $S_{v_n v_n}(e^{j\omega})$  is defined by (5).

The power spectrum defined here is equal in value to the power spectrum defined by Tjøstheim<sup>7</sup>, but Tjøstheim's definition has a different interpretation. Tjøstheim's power spectrum is defined to satisfy certain operator theoretic properties. Several other attempts have been made to define an appropriate power spectrum for nonstationary processes. These include an instantaneous power spectrum by Page<sup>8</sup>, the Wigner-Ville distribution<sup>9</sup>, and the evolutionary spectrum by Priestley<sup>10</sup>. For a summary of frequency-domain techniques for nonstationary processes see Treviño<sup>11</sup>.

In 1968 Loynes<sup>12</sup> listed the properties which a power spectrum for nonstationary processes should satisfy. Although all of the definitions mentioned partially satisfy this list, a power spectrum which satisfies all of these properties has not yet been discovered. Some of the properties proposed by Loynes which are satisfied by the power spectrum  $S_v(e^{j\omega}, n)$  defined by (8) are listed next.

**Proposition 4.** The power spectrum  $S_v(e^{j\omega}, n)$  defined by (8) is real valued and nonnegative with  $S_v(e^{j\omega}, n) = S_v(e^{-j\omega}, n)$  for all integers  $n$  and  $\omega \in \mathbb{R}$  (where  $\mathbb{R}$  is the set of real numbers).

**Proposition 5.** The instantaneous power in the process  $v(n)$  at index  $n$  is equal to the instantaneous power in the spectrum  $S_v(e^{j\omega}, n)$  at index  $n$ ; that is,

$$E[v(n)^2] = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_v(e^{j\omega}, n) d\omega$$

As Loynes pointed out, in generalizing any theory to cover a more general situation, an obvious requirement is that the new theory should be consistent with the earlier formulation. As shown below, this is the case for the power spectrum  $S_v(e^{j\omega}, n)$  defined by (8).

**Proposition 6.** If the signal model for the process  $v(n)$  is LTI then the process  $v(n)$  is WSS and the generalized power spectrum  $S_v(e^{j\omega}, n)$  reduces to the ordinary power spectrum  $S_v(e^{j\omega})$ .

Another of the requirements set forth by Loynes is that the power spectrum should transform simply when the process is transformed linearly. Consider the process  $y(n)$  formed by passing the process  $v(n)$ , with signal model defined by the frequency-response function  $G(e^{j\omega}, n)$ , through the linear filter with frequency-response function  $H(e^{j\omega}, n)$  as shown in Figure 3.

**Proposition 7.** The power spectrum of  $y(n)$  is given by

$$S_y(e^{j\omega}, n) = |G(e^{j\omega}, n)|^2 S_v(e^{j\omega}, n) \quad (9)$$

where

$$\Gamma(e^{j\omega}, n) = \sum_{i=-\infty}^n h(n, i) \frac{G(e^{j\omega}, i)}{G(e^{j\omega}, n)} e^{-j\omega(n-i)}$$

and  $h(n, m)$  is the impulse-response function of the filter with frequency-response function  $H(e^{j\omega}, n)$ .

Note that if the cascaded filters in Figure 3 are time invariant then  $\Gamma(e^{j\omega}, n) = H(e^{j\omega}, n) = H(e^{j\omega})$ . In this case (9) reduces to the well-known result  $S_y(e^{j\omega}) = |H(e^{j\omega})|^2 S_v(e^{j\omega})$  for the stationary case.

Although the relationship given by equation (9) is simple, using it to find  $S_y(e^{j\omega}, n)$  requires finding  $\Gamma(e^{j\omega}, n)$  which can be very difficult. Therefore, we consider the simplifying approximation given by the quantity  $\hat{S}_y(e^{j\omega}, n) = |H(e^{j\omega}, n)|^2 S_v(e^{j\omega}, n)$ . We will use this approximation in Section 5. We now conclude this section by investigating the accuracy of this approximation as a function of the rate of variation in  $v(n)$ .

Consider the quantity  $G(e^{j\omega}, e^{j\phi})$  defined by

$$G(e^{j\omega}, e^{j\phi}) = \sum_{n=-\infty}^{\infty} G(e^{j\omega}, n) e^{-j\phi n}$$

Clearly if the signal model  $G(e^{j\omega}, n)$  is time invariant, then the process  $v(n)$  is stationary and  $G(e^{j\omega}, e^{j\phi}) = G(e^{j\omega}, 0) \delta(\phi)$  for  $-\pi \leq \phi \leq \pi$ . Kailath<sup>13</sup> observes that the frequency variable  $\phi$  in  $G(e^{j\omega}, e^{j\phi})$  corresponds to the rate of variation in the signal model. It is reasonable to conclude that as the bandwidth of  $G(e^{j\omega}, e^{j\phi})$  in  $\phi$  increases so does the rate of variation in the signal model.

Assume that there exists finite constant values  $B_G$ ,  $B_H$ , and  $B_\phi$  such that

$$\begin{aligned} |G(e^{j\omega}, n)| &< B_G \\ |H(e^{j\omega}, n)| &< B_H \end{aligned} \quad (10)$$

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |\phi| |G(e^{j\omega}, e^{j\phi})| d\phi < B_\phi \quad (11)$$

for all integers  $n$  and  $\omega \in \mathbb{R}$ . Since it would appear reasonable to regard the smallest values of  $B_\phi$  satisfying (11) as corresponding to the bandwidth of  $G(e^{j\omega}, e^{j\phi})$  in  $\phi$ , it in turn also corresponds to the rate of variation for the processes  $v(n)$ .

It will also be assumed that there exists a finite constant value  $B_h$  such that

$$\sum_{i=0}^{\infty} |i| |h(n, n-i)| \leq B_h \quad (12)$$

for all integers  $n$ . Roughly speaking, the smallest value  $B_h$  for which (12) holds corresponds to the maximum duration of the impulse-response function  $h(n, n-i)$ . Now the following result can be stated.

**Proposition 8.** Let  $B_G$ ,  $B_H$ ,  $B_\phi$ , and  $B_h$  be given by (10), (11), and (12) respectively. Then with  $B_G$ ,  $B_H$ , and  $B_h$  fixed,  $|S_y(e^{j\omega}, n) - \hat{S}_y(e^{j\omega}, n)| \rightarrow 0$  uniformly in  $\omega$  and  $n$  as  $|B_\phi| \rightarrow 0$ .

**Proof.** First define  $\Psi_G(e^{j\omega}, n)$  by

$$\Psi_G(e^{j\omega}, n) = \Gamma_G(e^{j\omega}, n) G(e^{j\omega}, n) - H(e^{j\omega}, n) G(e^{j\omega}, n)$$

Now it is easily verified that

$$\Gamma_G(e^{j\omega}, n) G(e^{j\omega}, n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} H(e^{j(\phi+\omega)}, n) G(e^{j\omega}, e^{j\phi}) e^{j\phi n} d\phi \quad (13)$$

Using the Mean-Value Theorem (13) can be expressed as

$$\Gamma_G(e^{j\omega}, n) G(e^{j\omega}, n) = H(e^{j\omega}, n) G(e^{j\omega}, n) + \frac{1}{2\pi} \int_{-\pi}^{\pi} \phi \frac{d}{d\phi} H(e^{j(\omega+\gamma\phi)}, n) G(e^{j\omega}, e^{j\phi}) e^{j\phi n} d\phi$$

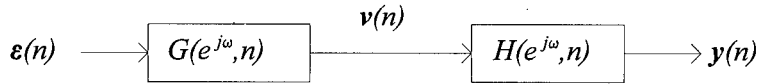
where  $0 \leq \gamma \leq 1$ . Using the bounds given by (11) and (12), it follows that

$$|\Psi_G(e^{j\omega}, n)| < B_h B_\phi \quad (14)$$

From (14) it follows easily that

$$|S_y(e^{j\omega}, n) - \hat{S}_y(e^{j\omega}, n)| < (B_h B_\phi)^2 + 2B_h B_G B_h B_\phi$$

■



**Figure 3.** Block diagram of  $G(e^{j\omega}, n)$  cascaded with  $H(e^{j\omega}, n)$ .

The power spectrum is used to give the optimal solution to the clairvoyant detection problem described in the next section, and in Section 5 it is used to give a suboptimal solution to the detection problem described in Section 2.

#### 4. Clairvoyant Detection

Consider the clairvoyant detection problem depicted in Figure 4 where we somehow are given the family of processes  $\{r_k(n, n_d) = s(n - n_d) + v_k(n)\}$  associated with the measured process  $r(n, n_d)$ . For a fixed integer  $k$ , the received signal  $r_k(n, n_d)$  is passed through the filter  $h_k(n)$  to generate the response  $y_k(n, n_d)$  given by

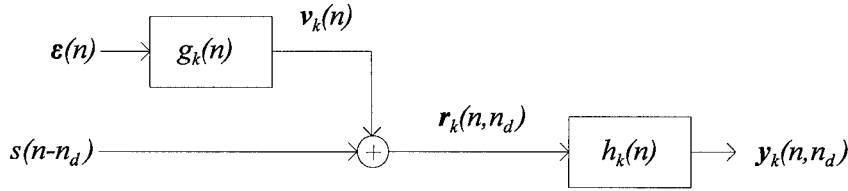
$$y_k(n, n_d) = \sum_{m=-\infty}^n h_k(n-m) r_k(m, n_d) = y_{s_k}(n, n_d) + y_{v_k}(n, n_d) \quad (15)$$

The objective in this problem is to design the family of time-invariant filters  $\{h_{MF_k}(n)\}$  such that with  $h_k(n) = h_{MF_k}(n)$ , the family of signal-to-noise ratios  $R_k(n, n_d)$  defined by

$$R_k(n, n_d) = \frac{|y_{s_k}(n, n_d)|^2}{r_{y_{v_k}}(n, n)} \quad (16)$$

are maximized at the specific index  $n=n_p=n_d+n_0$  where  $n_0$  is a design parameter and

$$r_{y_{v_k}}(n, n) = E[y_{v_k}(n)y_{v_k}(n)]$$



**Figure 4.** Block diagram of clairvoyant detection problem.

The optimal solution to this clairvoyant detection problem is given by the following theorem.

**Theorem 9.** With  $h_k(n)=h_{MF_k}(n)$  the family of time-invariant filters  $h_{MF_k}(n)$  which maximize the family of signal-to-noise ratios  $R_k(n, n_d)$  at the index  $n=n_p=n_d+n_0$  is given by

$$H_{MF_k}(e^{j\omega}) = \frac{S^*(e^{j\omega})e^{-j\omega n_0}}{S_v(e^{j\omega}, k)}$$

where  $S(e^{j\omega})$  is the Fourier transform of  $s(n)$  and

$$H_{MF_k}(e^{j\omega}) = \sum_{n=-\infty}^{\infty} h_{MF_k}(n) e^{-j\omega n}$$

**Proof.** Write (16) in the modified form

$$R_k(n, n_d) = \frac{\left| \int_{-\pi}^{\pi} H_{MF_k}(e^{j\omega}) \sqrt{S_v(e^{j\omega}, k)} \frac{S(e^{j\omega}) e^{-j\omega n_d}}{\sqrt{S_v(e^{j\omega}, k)}} e^{j\omega n} d\omega \right|^2}{2\pi \int_{-\pi}^{\pi} |H_{MF_k}(e^{j\omega})|^2 S_v(e^{j\omega}, k) d\omega}$$

Employing the Scharz inequality gives

$$R_k(n, n_d) \leq \frac{\int_{-\pi}^{\pi} H_{MF_k}(e^{j\omega}) S_v(e^{j\omega}, k) d\omega \int_{-\pi}^{\pi} \frac{|S(e^{j\omega})|^2}{S_v(e^{j\omega}, k)} d\omega}{2\pi \int_{-\pi}^{\pi} |H_{MF_k}(e^{j\omega})|^2 S_v(e^{j\omega}, k) d\omega}$$

with equality holding at the index  $n=n_p=n_d+n_0$  if

$$H_{MF_k}(e^{j\omega}) = \frac{S^*(e^{j\omega}) e^{-j\omega n_0}}{S_v(e^{j\omega}, k)}$$

■

## 5. Suboptimal Detection

Recall that the detection problem of Figure 2 calls for finding the time-varying filter  $h_{MF}(n, m)$  such that with  $h(k, k-n) = h_{MF}(k, k-n)$ , the signal-to-noise ratio  $R(n, n_d)$  is maximized. Consider the suboptimal detector defined in terms of the optimal clairvoyant detector according to  $h(k, k-n) = h_{MF_k}(n)$  with response  $\hat{y}(n, n_d)$  given by

$$\hat{y}(n, n_d) = \sum_{m=-\infty}^n h_{MF_k}(n-m) r(m, n_d) \quad (17)$$

The following theorem states that under slowly varying conditions, the signal-to-noise ratio  $R(n, n_d)$  resulting from  $h(k, k-n) = h_{MF_k}(n)$  will be close to the *minimum* signal-to-noise ratio resulting from  $h(k, k-n) = h_{MF}(k, k-n)$ . The statement of this theorem requires the existence of fixed finite values  $b_\Gamma$ ,  $b_G$ ,  $b_H$ , and  $B_{y_s}$  such that

$$|\Gamma(e^{j\omega}, n)| > b_\Gamma > 0 \quad |G(e^{j\omega}, n)| > b_G > 0 \quad |H(e^{j\omega}, n)| > b_H > 0 \quad (18)$$

for all integers  $n$  and  $\omega \in \mathbb{R}$  and

$$|y_s(n, n_d)|^2 < B_{y_s} \quad (19)$$

for all integers  $n$  and all positive integers  $n_d$ .

**Theorem 10.** Consider the detection problems diagrammed in Figure 2 and Figure 4. Let  $b_\Gamma$ ,  $b_G$ ,  $b_H$ ,  $B_{y_s}$ ,  $B_G$ ,  $B_H$ ,  $B_\phi$ , and  $B_h$  be given by (18), (19), (10), (11), and (12) respectively. Then with  $b_\Gamma$ ,  $b_G$ ,  $b_H$ ,  $B_{y_s}$ ,  $B_G$ ,  $B_H$ , and  $B_h$  fixed and  $h(k, k-n) = h_k(n)$ ,  $|R(n, n_d) - R_n(n, n_d)| \rightarrow 0$  uniformly in  $\omega$  and  $n$  as  $|B_\phi| \rightarrow 0$ .

**Proof.** First observe that  $H(e^{j\omega}, k) = H_k(e^{j\omega})$ . Next write

$$|R(n, n_d) - R_n(n, n_d)| = \left| \frac{y_s(n, n_d)}{\frac{1}{2\pi} \int_{-\pi}^{\pi} S_{y_s}(e^{j\omega}, n) d\omega} - \frac{y_s(n, n_d)}{\frac{1}{2\pi} \int_{-\pi}^{\pi} |H(e^{j\omega}, n)|^2 S_v(e^{j\omega}, n) d\omega} \right|$$

Then cross multiply to form a common denominator



$$|R(n, n_d) - R_n(n, n_d)| = \left| \frac{2\pi y_s(n, n_d) \int_{-\pi}^{\pi} |H(e^{j\omega}, n)|^2 S_v(e^{j\omega}, n) - S_{y_v}(e^{j\omega}, n) d\omega}{\int_{-\pi}^{\pi} S_{y_v}(e^{j\omega}, n) d\omega \int_{-\pi}^{\pi} |H(e^{j\omega}, n)|^2 S_v(e^{j\omega}, n) d\omega} \right|$$

It follows from the given inequalities and Proposition 8 that

$$|R(n, n_d) - R_n(n, n_d)| < \frac{B_{y_s} (B_h B_\phi)^2 + 2 B_H B_G B_h B_\phi}{(b_\Gamma b_\phi^2 b_H)^2} \quad (19)$$

■

With  $h(n, n-m) = h_n(m)$  Theorem 10 states that if the rate of variation in  $v(n)$  is sufficiently small, then the functions  $R(n, n_d)$  and  $R_n(n, n_d)$  are close. An explicit bound on the difference  $|R(n, n_d) - R_n(n, n_d)|$  is given by equation (19). The point is that when the signal-to-noise ratio  $R(n, n_d)$  resulting from  $h_k(m) = h_{MF_k}(m)$  is close to the *maximum* signal-to-noise ratio resulting from  $h(n, n-m) = h_{MF}(n, n-m)$ , then the detector response  $\hat{y}(n)$  defined by (17) and generated using  $h(m, k) = h_{MF_k}(m)$  is a "viable" suboptimal solution.

## 6. Numerical Example

In this section we present a numerical example which compares the response of the suboptimal detector defined by (17) with that of the optimal clairvoyant detector defined by (15).

In this example we wish to detect the signal  $s(n) = \sin(n\pi/4)$   $n=0, 2, \dots, 31$  in the received signal  $r(n, n_d) = s(n - n_d) + v(n)$ . The signal  $s(n - n_d)$  for  $n_d = 128$  is shown in Figure 5. The nonstationary interference  $v(n)$  has signal model given by

$$G(e^{j\omega}, n) = \frac{b(n)}{1 + a_1(n) e^{-j\omega} + a_2(n) e^{-2j\omega}}$$

where the numerator  $b(n)$  is chosen to normalize the process power over time, and the denominator polynomial has roots  $Me^{\pm j\omega(n)}$  where  $M=0.9$  and

$$\omega(n) = \begin{cases} \frac{\pi}{4}, & n \leq 0 \\ \frac{\pi}{4} + \left( \frac{3\pi}{4} - \frac{\pi}{4} \right) \frac{n}{256}, & 1 \leq n \leq 256 \\ \frac{3\pi}{4}, & n \geq 257 \end{cases}$$

A sample realization of the received signal  $r(n, n_d)$  is shown in Figure 6. Observe that the frequency content of the sample realization in Figure 6 changes such that the initial frequency is lower than the final frequency.

Figure 7 shows the detector response if one treats the noise as stationary and white. The optimal clairvoyant detector response  $y(n, n_d) = y_n(n, n_d)$  defined by (15) is shown in Figure 8, and the suboptimal detector response  $\hat{y}(n, n_d)$  defined by (17) is shown in Figure 9. Comparing Figures 7, 8, and 9 one finds that treating the interference as stationary and white is inadequate for detection purposes, but one also finds that the suboptimal detector provides a response that is both adequate and nearly identical to that of the optimal clairvoyant detector.

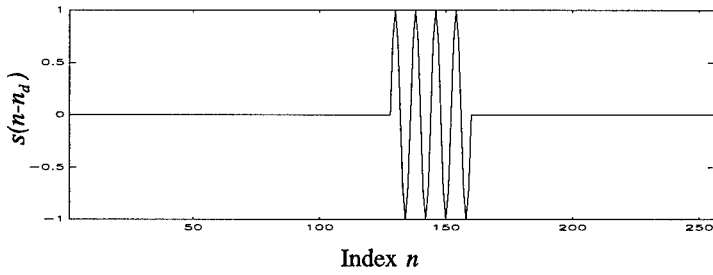


Figure 5. Received signal.

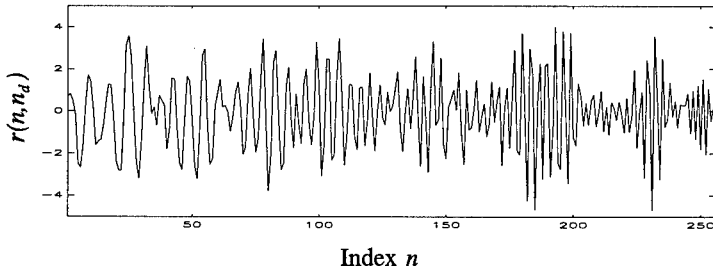
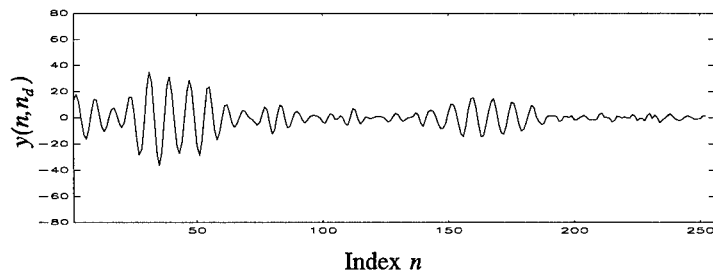
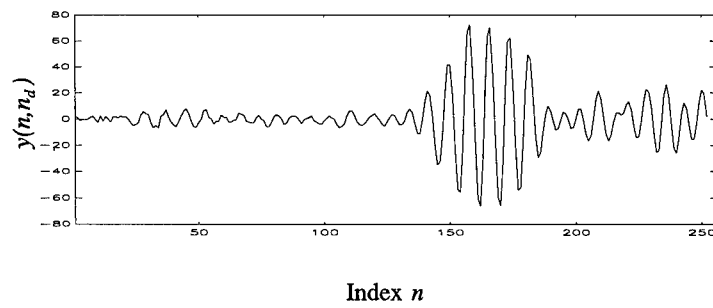


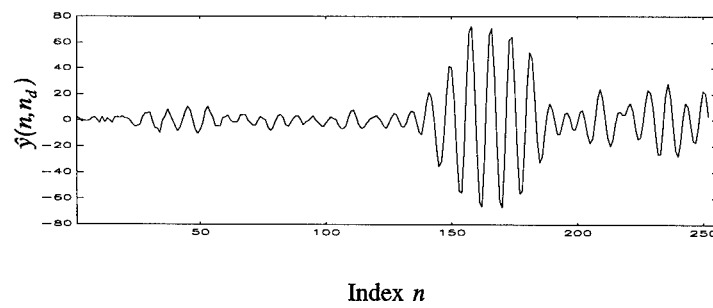
Figure 6. Received noisy signal.



**Figure 7.** White noise detector response.



**Figure 8.** Optimal clairvoyant response.



**Figure 9.** Suboptimal response.

## 7. Conclusions

We showed that associated with every PND process is a family of jointly wide-sense stationary processes. This fundamental interpretation has some interesting theoretical implications including a definition for the power spectrum of a nonstationary process. This interpretation was also used to formulate a clairvoyant detection problem with optimal solution given in terms of the power spectrum. This solution was applied to the nonstationary detection problem, and its performance was examined in terms of the process rate of variation. It was shown that this solution provides nearly optimal performance when the rate of variation is sufficiently small. Future work includes developing techniques to treat the detection problem for the case of rapidly varying nonstationary processes.

## References

1. G.L. Turin, *IRE Trans. Inform. Theory*, **6** (1960), 311-329.
2. R. Price and P.E. Green, *Proc. IRE*, **46** (1958), 555-570.
3. J.W. Modestino and K.A. Melendez, *Proc. Signal and Data Processing of Small Targets*, SPIE **1954** (1993), 155-166.
4. H. Cramér, *Proc. Berkeley Symp. Math. Stat. and Probab.*, **2** (1961), 57-78.
5. J.A. Sills, *Ph.D. Thesis*, Georgia Institute of Technology, (1995).
6. L.A. Zadeh, *Proc. I.R.E.*, **38** (1950), 291-299.
7. D. Tjøstheim, *Adv. Appl. Prob.*, **8** (1976), 831-846.
8. C.H. Page, *J. Appl. Phys.*, **23** (1952) 103-106.
9. L. Cohen, *Proc. IEEE*, **77** (1989) 941-989.
10. M.B. Priestley, *Spectral Analysis and Time Series*, (Academic Press, 1992).
11. G. Treviño, *Proc. First Workshop on Nonstationary Stochastic Processes and their Applications*, (1991) 48-61.
12. R.M. Loynes, *J. Roy. Stat.*, **B30** (1968), 1-30.
13. T. Kailath, *MIT Research Lab. of Electronics, Tech. Report 352*, (1959).

---

## INCONSISTENCIES IN PARAMETERS ESTIMATED FROM IMPULSIVE NOISE

Robert D. Pierce

*Naval Surface Warfare Center-Carderock Division*  
*Code 7340 (7029)*  
*Bethesda, MD 20084-5000*  
rpierce@oasys.dt.navy.mil

### ABSTRACT

Impulsive noise described by an alpha-stable distribution has second-order as well as higher-order moments that are infinite. For this noise, estimates of conventional parameters such as variance and power spectra are not consistent and can appear to be non-stationary. The particular type of impulsive noise examined consists of complex noise samples having a bivariate isotropic alpha-stable distribution. The complex-Gaussian distribution is a special case where the envelope is Rayleigh distributed and, only for this case, all moments exist and are consistent. When the noise is highly impulsive, even the mean or first moment may not exist. Examples of moderately impulsive noise include sea clutter, the radar returns from ocean waves, and especially high-resolution, horizontally-polarized returns. Spectral analysis of this noise, such as Doppler processing, can retain the impulsive nature of the noise and make the spectral estimates inconsistent. The concept of positive and negative lower-order moments of the envelope is used to provide robust methods for obtaining consistent estimates in the presence of impulsive noise.

## 1. Introduction

### 1.1 General

From the viewpoint of the measurement, detection and characterization of signals and noise, the most commonly assumed probability distribution function is probably the Gaussian distribution. This would include the many Gaussian derived distributions such as the Rayleigh, Chi-squared, noncentral Chi-squared, Student t, F, etc. The statistics that describe observed physical phenomena are defined by this assumption, and the signal processing methods to obtain parameter estimates and test statistics are derived under this assumption. The Gaussian assumption can generally be justified using the central limit theorem<sup>1</sup> where the physical phenomena being studied consists of sums of identically distributed, statistically independent events with finite means and variances. The "identically distributed" requirement can be dropped if moments higher than the variance exist. The Gaussian assumption, however, is inappropriate when the means and variances of the independent events are infinite.

When these moments are infinite, the generalized central limit theorem<sup>2,3</sup> states that the sums converge to the alpha-stable distribution of which the Gaussian distribution is a subset. These distributions can model physical phenomena that are impulsive or spiky in nature. Such phenomena, for example, can include radio static from lightning storms, as well as high-resolution radar returns from sea clutter. Problems begin when these data are processed using methods based on conventional, second-order statistics under the Gaussian assumption.

If data with an alpha-stable distribution are processed using ordinary second-order methods such as variance, correlation or power spectrum estimation, then the results can appear to be non-robust, inconsistent or nonstationary. If the impulsive nature of the data is recognized, then ad-hoc processing methods such as clipping or the removal of "outliers" are used. This also results in the use of other statistics such as the median or the geometric mean that can limit the description or characterization of the data. Empirically derived distributions such as log-normal or Weibull are often fitted to the data.

An example of inconsistent mean-square estimates is given in Fig. 1. The top trace shows the envelope of bin #16 from a sea clutter Doppler spectrum (such as from one velocity bin of a Doppler spectrum) as a function of sample number (proportional to time). The bottom traces show running averages for estimates of the root-mean-square (RMS, square-root of the second moment) and the fourth power of the negative 0.25-order moment. These two statistics are proportional to amplitude. The running average for the RMS shows the inconsistent behavior associated with infinite variance. In contrast, the negative 0.25 order moment stabilizes and shows little fluctuation.

When the signals and noise from a physical phenomena are recognized to be alpha-stable distributed, the measurement, detection and characterization of this phenomena can be accomplished using methods that are robust and give consistent, generalized results. This paper presents examples where the theory is applied to radar returns from sea clutter for detection, estimation of moments and distribution parameters, and Doppler spectrum analysis.

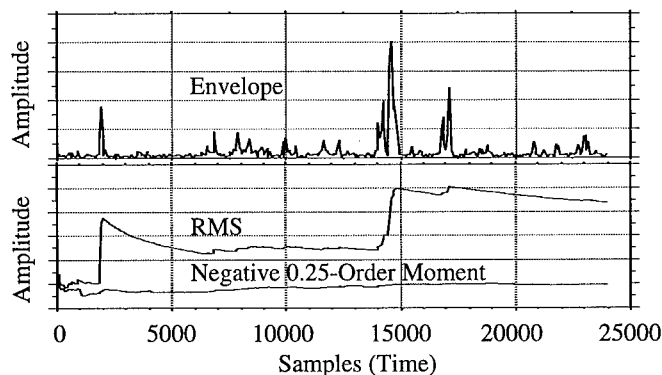


Fig. 1. Example of Inconsistent RMS Estimate

### 1.2 Alpha-Stable Distributions

The study of alpha-stable distributions and their application to signal processing has grown significantly over the past several years. A tutorial review of the basic characteristics of stable distributions and stable signal processing was given in 1993 by Shao and Nikias<sup>1</sup>. This review was followed with a book by Nikias and Shao<sup>2</sup>, published in 1995, that provides an introduction to stable distributions and processes and that emphasizes the application of alpha-stable concepts to signal processing. In the field of stochastic modeling, Samorodnitsky and Taqqu<sup>4</sup> published a book in 1994 that targets researchers in probability as well as applied probability and statistics.

For the alpha-stable distribution, the parameter alpha,  $\alpha$ , the characteristic exponent, varies over  $0 < \alpha \leq 2$ . The stable distribution includes the Gaussian when  $\alpha = 2$ . As  $\alpha$  becomes smaller than 2, the random process becomes more impulsive and more non-Gaussian in nature—the tails of the stable distribution become thicker. Hence, the stable distribution is an attractive choice for modeling signals and noise with an impulsive nature. For  $\alpha < 2$ , the random process has infinite variance (including moments higher than second-order); however, a range of fractional lower-order and negative-order moments exist that are consistent.

From the generalized central limit theorem, the stable distribution is the only limiting distribution for sums of independent and identically distributed random variables. In a quote from Shao and Nikias<sup>2</sup>: “If an observed signal or noise can be thought of as the sum or results of a large number of independent and identically distributed effects, then the Generalized Central Limit Theorem suggests that a stable model may be appropriate.” This is related to the *stability property* quoted from Shao and Nikias<sup>2</sup>: “the sum of two independent stable random variables with the same characteristic exponent is again stable and has the same characteristic exponent.”

### 1.3 Radar Returns from Sea Clutter

Stable distributions can model radar returns from sea clutter, the radar reflections from ocean waves. Radar returns are generally more spiky when using horizontal polarization for transmit and receive, and when using high spatial resolution. The time-varying clutter tends to prevent radar detection of targets; the spiky nature of the clutter tends to raise the false-alarm rate. Knowledge of the distribution can result in a detector designs that are tailored to the type of clutter. In Nohara and Haykin<sup>5</sup> the targets of interest are small fragments of icebergs.

As discussed by Nohara and Haykin<sup>5</sup> and Armstrong and Griffiths<sup>6</sup>, the K-distribution is currently a widely accepted model for representing the amplitude or envelope statistics of sea clutter. The distribution is based on the assumption that the radar return from a region consists of a sum of independent returns (speckle) that vary in intensity with time. The I and Q (in-phase and quadrature) returns from sea clutter are described<sup>5</sup> as a complex Gaussian process

(short decorrelation time) modulated by a second random process (long decorrelation time) that is chi-distributed. The modulation results in spiky radar returns.

Field data, however, suggest that the K-distribution does not fully explain observed sea clutter distributions. An example using horizontal (H-pol) polarized sea clutter is presented later in this paper. The K-distribution is shown to match the main body of the distribution; however, it fails to match the tail. The alpha-stable distribution is shown to match both.

## 2. Alpha-Stable Envelope and Its Moments

The first applications of the alpha-stable distribution are related to the envelope and its moments. The probability density function for the envelope of bivariate isotropic stable distributed noise<sup>3</sup> is

$$f(z) = z \int_0^\infty s e^{-\gamma s^\alpha} J_0(sz) ds \quad (1)$$

where the noise envelope,  $z$ , is given by

$$z = \sqrt{z_I^2 + z_Q^2} \quad (2)$$

and the in-phase and quadrature components  $z_I$  and  $z_Q$  are jointly isotropic alpha-stable with dispersion,  $\gamma$ . The  $p^{\text{th}}$  moment is defined by

$$m_p = E[z^p] = \int_0^\infty z^p p(z) dz \quad (3)$$

For the  $M^{\text{th}}$  variate (for the bivariate case,  $M = 2$ ), the moment<sup>3</sup> is

$$m_p = 2^p \frac{\Gamma\left(\frac{p+M}{2}\right) \Gamma\left(1-\frac{p}{\alpha}\right)}{\Gamma\left(\frac{M}{2}\right) \Gamma\left(1-\frac{p}{\alpha}\right)} \gamma^{\frac{p}{\alpha}} \quad (4)$$

For  $\gamma > 0$  and  $0 < \alpha < 2$ , then  $-M < p < \alpha$ . For the Gaussian case,  $\alpha = 2$ ,  $-M < p < \infty$ , and the variance is  $\sigma^2 = 2\gamma$ .

The moment that appears to be particularly useful for normalizing the envelopes of stable bivariate distributions is the negative first-order moment (NFM)

$$m_{-1} = \frac{\Gamma(1/\alpha)}{\alpha \gamma^{1/\alpha}} \quad (5)$$

which can be calculated from estimates of  $\alpha$  and  $\gamma$ , or the NFM can be estimated from

$$\hat{m}_{-1} = \frac{1}{N} \sum_{k=1}^N \frac{1}{z_k} \quad (6)$$

The NFM estimator is unbiased; however, as shown later, its variance is infinite. Since the envelope density function is scaled<sup>3</sup> by  $z/\gamma^{1/\alpha}$ , the envelope samples can be normalized in amplitude by multiplying the samples by the negative first-order moment estimate. This normalization method will be used in this paper.



The envelope density function given by Eq. (1) only has a closed form solution for the Gaussian or Rayleigh ( $\alpha = 2$ ) and the isotropic Cauchy ( $\alpha = 1$ ) density functions. The Rayleigh density<sup>3</sup> is

$$f(z) = \frac{z}{2\gamma} e^{-z^2/4\gamma} \quad (7)$$

and the isotropic Cauchy density<sup>3</sup> is

$$f(z) = \frac{z\gamma}{(z^2 + \gamma^2)^{3/2}} \quad (8)$$

Otherwise the density function is calculated by numerical integration of Eq. (1). Graphs of the density function normalized by the NFM are given in Fig. 2.

Methods for estimating  $\alpha$  are presented in the literature<sup>2,3</sup>; however, for the envelope, the graphs of the density function in Fig. 2 indicate that a simple approach would be to count how often the envelope exceeds a specified level. Since a "pivot" exists at a normalized amplitude of 3.1, setting the threshold at this level would give the greatest variation in the number of occurrences for different  $\alpha$ . Fig. 3 gives the probability of exceeding the normalized amplitude of 3.1 and 2.5. These probabilities were obtained by numerical integration of the density functions. A two-pass operation is required on the data: the NFM is estimated first, followed by a second pass to count the number of times the levels are exceeded. If sufficient

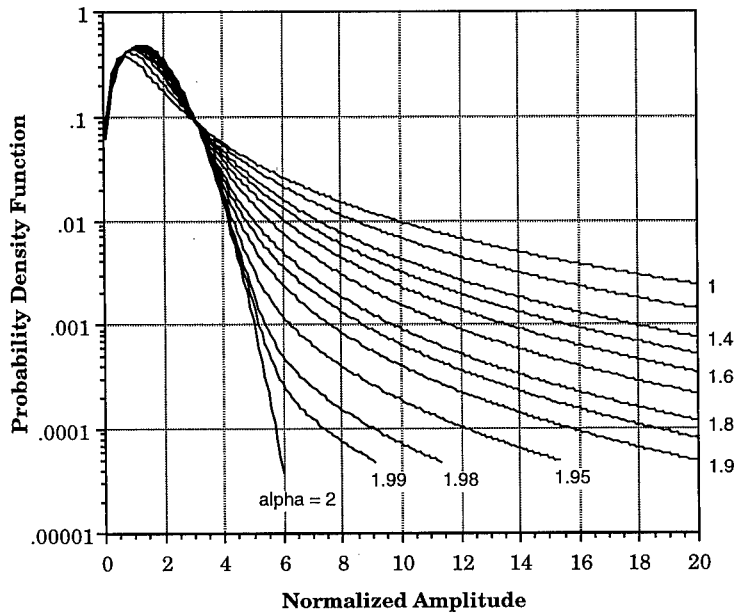


Fig. 2. Envelope Distribution for Bivariate Isotropic Stable Distributed Noise

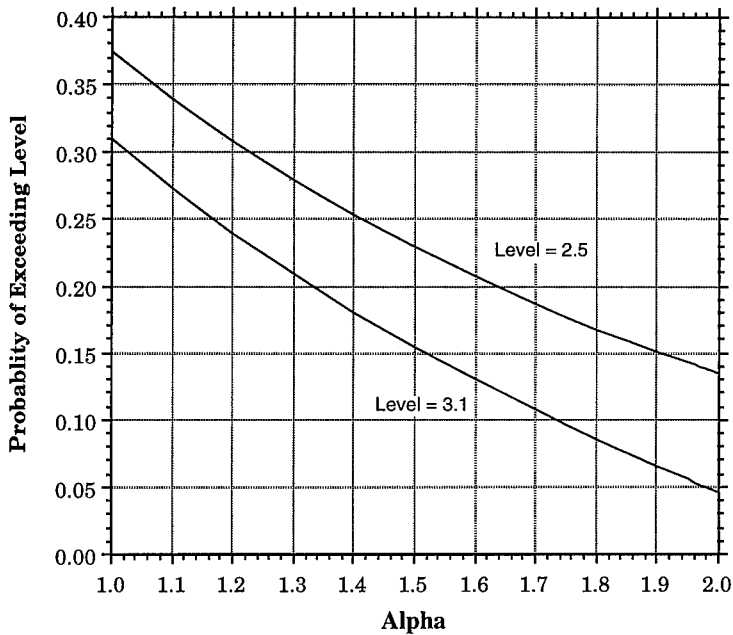


Fig. 3. Probability of Exceeding Normalized Envelope Levels of 3.1 and 2.5

samples exists, then comparison of a histogram to the calculated distribution is still necessary to verify the fit to the tail of the distribution.

Next, the median and mode are numerically calculated from the envelope density and normalized by the NFM. As shown in Fig. 3, the mode or peak of the distribution (the most probable value) varies almost linearly as alpha varies from 1 to 2. Of greater interest is the variation of the median with alpha. Since the median is nearly constant as alpha varies from about 1.4 to 2, the NFM is an close estimator of the median and vice versa. The median equals 1.48 times the NFM. The circles at the end points indicate median values calculated from Eq. (7) and Eq. (8).

If second-order statistics are used with alpha-stable distributed data where saturation or clipping occurs, then the second-order statistics can still give arbitrary results. This can easily occur, for example, in a radar system where the radio frequency (RF) amplifiers and/or the analog-to-digital (A/D) converters are driven into saturation by large spikes from the clutter returns. To show the variation in root-mean-square (RMS) as a function of alpha and clipping level, the second moment (mean-square) was calculated by numerically integrating Eq. (1). When the saturation level was reached, the density function was replaced by a delta function that represented the remaining area under the density function beyond the clipping level. Since the tail no longer extends to infinity, the second moment is finite. The results of this integration are presented in Fig. 5. The RMS is normalized by the NFM (assumed to be

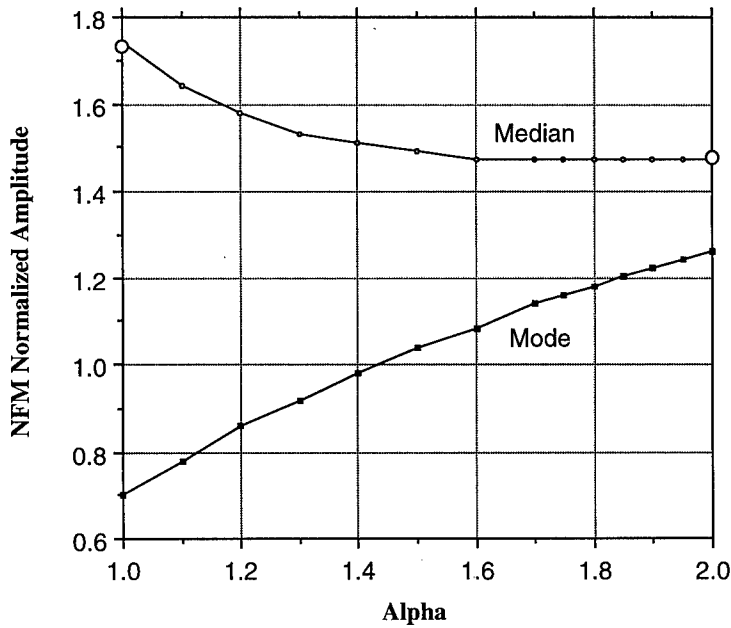


Fig. 4. Median and Mode for Envelope of Bivariate Isotropic Stable Distribution

known). The clipping level is likewise normalized by the NFM. Alpha was varied from 1 to 2 and the clipping level was varied from 1 to 20. At alpha equal to 2 and no clipping, the RMS is  $\sqrt{\pi}$ . For alpha equal to 2, clipping levels above 4 have very little affect on the RMS. However, for alpha equal to 1, the RMS varies from 2.5 to 6.5 as the clipping level varies from 4 to 20. The results are highly sensitive to clipping level and alpha.

In contrast, when the negative first-order moment is calculated using numerical integration and clipping, the NFM calculation normalized by the known NFM is constant at unity for all clipping levels from 3 to 20. The NFM, like the median, is not very affected by saturation.

### 3. Sea Clutter as a Stable Process

#### 3.1 Fit to Histogram

Using the NFM to normalize the histograms and using the probability of exceeding method to estimate  $\alpha$ , an example of fitting sea clutter histograms to the stable envelope distribution is given in Fig. 5 for horizontally polarized (H-Pol) sea clutter returns. The comparison includes the K-distribution that was also normalized by the negative first-order moment. The sea clutter data were taken in a sea state 3 with an X-band radar at 8° look-down

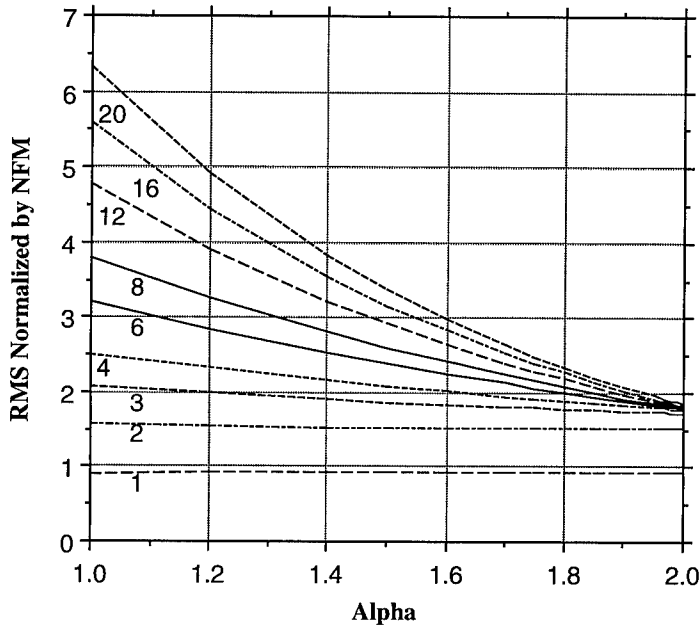


Fig. 5. RMS at Different Normalized Clipping Levels

angle with a spatial resolution of 1.52 m. This represents a nominal sea condition. The clutter data in Fig. 5 had an exceeding probability of 0.109 which is  $\alpha = 1.75$ . The K-distribution was fitted using the shape parameters  $\nu=4.0$ . The K-distribution closely matches the main body of the histogram up to a normalized amplitude of 4. Beyond this amplitude, the K-distribution diverges significantly and does not match the tail of the histogram. The stable distribution gives a much closer match to the entire histogram; however, the stable distribution does overestimate the tail. The stable distribution does a better job of modelling the impulsive nature of the H-Pol clutter.

### 3.2 Envelope Detectors

The performance of several envelope detectors are examined using the H-Pol sea clutter as noise or interference. The target is simulated as a sinewave with constant amplitude,  $A$ , with unknown, random phase,  $\theta_n$ , added to the I and Q signal. For radar, this is equivalent to a non-fading point target where from one sample to the next the phase varies randomly with a uniform distribution over the interval  $[0, 2\pi)$ . The envelope of this sum is the input to the

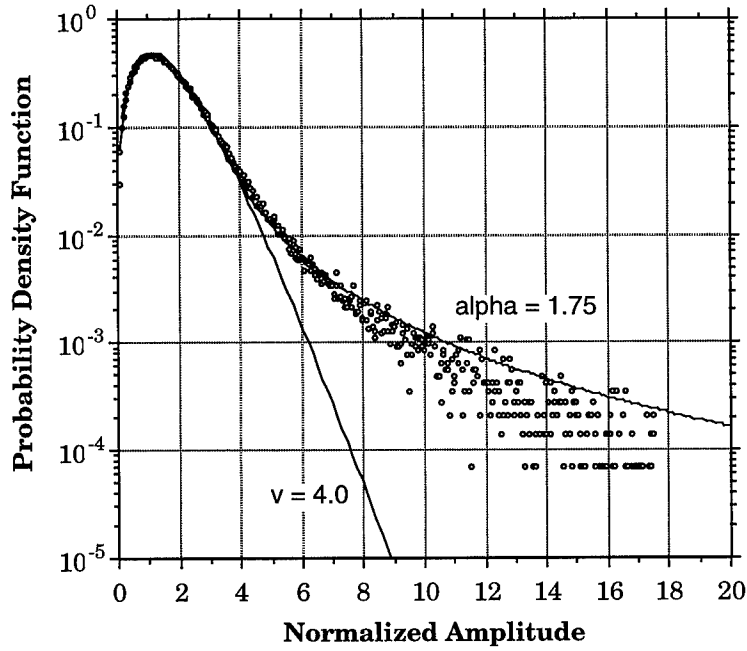


Fig. 6. Envelope Distribution for H-Pol Sea Clutter with 1.52 m resolution

detectors. The hypotheses are

$$H_1 : z_n = \sqrt{(z_{In} + A \cos \theta_n)^2 + (z_{Qn} + A \sin \theta_n)^2} \quad (9)$$

$$H_0 : z_n = \sqrt{z_{In}^2 + z_{Qn}^2}$$

Except for normalization by the NFM, the test statistic for the envelope detector is given by Whalen<sup>1</sup>

$$q_{env} = \frac{m_{-1}}{N} \sum_{n=1}^N z_n \quad (10)$$

The negative first-order moment (NFM) detector has a test statistic that is the reciprocal of the NFM estimator

$$q_{NFM} = \frac{m_{-1}}{\frac{1}{N} \sum_{n=1}^N \frac{1}{z_n}} \quad (11)$$

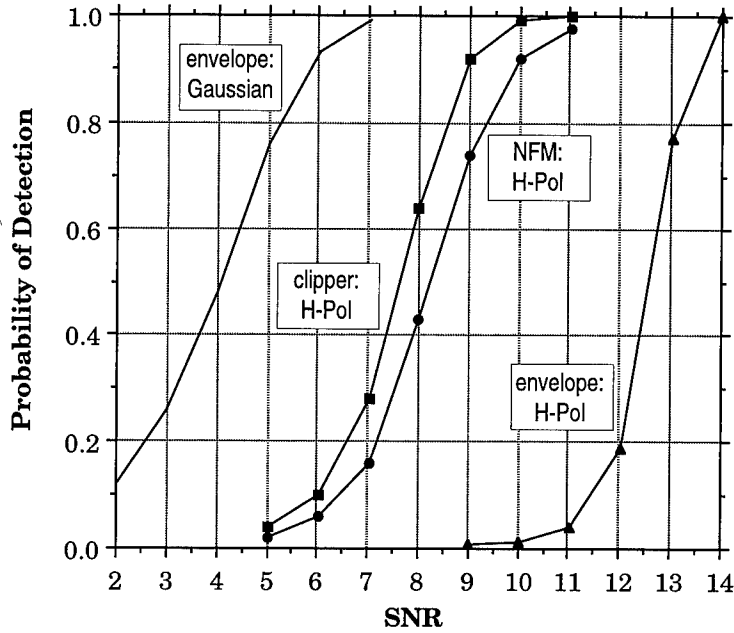


Fig. 7. Probability of Detection Curves for Different Detectors at PFA =  $10^{-4}$

The clipper is identical to the envelope detector with the envelope clipped at level  $C$ .

$$q_{clip} = \frac{m_{-1}}{N} \sum_{n=1}^N \begin{cases} z_n, & z_n < C \\ C, & z_n \geq C \end{cases} \quad (12)$$

Usually the detector is normalized by the noise-only standard deviation; however, since this moment does not exist, the test statistics are normalized by the negative first-order moment,  $m_{-1}$ , for the noise-only case. Constant False Alarm (CFAR) schemes would be based on estimates of this moment.

The results from applying these detectors to the H-Pol sea clutter are shown in Fig. 7 for the case of  $N = 32$  independent samples and a probability of false alarm (PFA) of  $10^{-4}$ . The signal-to-noise ratio (SNR) is defined as

$$SNR = 20 \log(A m_{-1}). \quad (13)$$

For Gaussian distributed noise,  $m_{-1}$  is related to standard deviation,  $\sigma$ , by

$$m_{-1} \sigma = \sqrt{\pi} \quad (14)$$

which is about 5 dB. In Fig. 7, the envelope detector in Gaussian noise has a probability of detection (PD) of 0.5 at 4 dB for PFA of  $10^{-4}$ . Because of the 5 dB difference in the definition of SNR, this corresponds to -1 dB for the curves given in Whalen<sup>1</sup> where the test statistic is normalized by standard deviation. The envelope detector in H-Pol clutter has PD=0.5 at 12.5

dB, so the envelope detector suffers an 8.5 dB loss in this H-Pol clutter (If the quadratic detector were used, then the loss would be closer to 12 dB.) The NFM detector shows a 4.5 dB improvement over the envelope detector while the clipper shows a 5 dB improvement. The clipper level was set to

$$C = 3/m_{-1} \quad (15)$$

For Gaussian noise, this corresponds to  $1.69\sigma$ . The performance of this detector is sensitive to small changes in clipper level. Because of the clipping, this detector may not work at very low false alarm rates. The NFM detector does not have these two problems, but it does have slightly worse performance.

#### 4. Generalized Cauchy Envelope Distribution

Closed form solutions for the negative first-order moment of a sinewave plus noise exist for the Gaussian and Cauchy distributed noise. The results for Gaussian distributed noise are given by Whalen. Similar results for Cauchy distributed noise are given below.

The envelope of a sinewave plus Cauchy distributed noise is derived from the bivariate isotropic Cauchy distribution<sup>3</sup>

$$f(z_I, z_Q) = \frac{\gamma}{2\pi (z_I^2 + z_Q^2 + \gamma^2)^{3/2}} \quad (16)$$

where  $\gamma$  is the dispersion. For the  $n^{\text{th}}$  sample, the transformation is

$$\begin{aligned} z_{In} &= z_n \cos \phi_n + A_n \cos \theta_n \\ z_{Qn} &= z_n \sin \phi_n + A_n \sin \theta_n \end{aligned} \quad (17)$$

where  $z_n$  is the envelope of the Cauchy noise with phase,  $\phi_n$ , uniformly distributed over  $[0, 2\pi)$ . The sinewave amplitude,  $A_n$ , and phase,  $\theta_n$ , can vary from sample to sample. The Jacobian is  $z_n$ , which gives (dropping the explicit sampling dependence)

$$f(z, \phi | A, \theta) = \frac{z \gamma}{2\pi (z^2 + A^2 + \gamma^2 + 2zA \cos(\phi - \theta))^{3/2}} \quad (18)$$

Next,

$$f(z | A, \theta) = \int_0^{2\pi} f(z, \phi | A, \theta) d\phi \quad (19)$$

so

$$f(z | A, \theta) = \frac{2 \gamma z}{\pi (a - b)(a + b)^{1/2}} E\left(\sqrt{2b/(a + b)}\right) \quad (20)$$

where  $a = z^2 + A^2 + \gamma^2$  and  $b = 2zA$  and  $E(\bullet)$  is the complete elliptic integral of the 2nd kind. The density is independent of  $\theta$ , so the density function of the envelope of a sinewave plus narrowband Cauchy noise is

$$f(z | A) = \frac{2 \gamma z}{\pi (a - b)(a + b)^{1/2}} E\left(\sqrt{2b/(a + b)}\right) \quad (21)$$

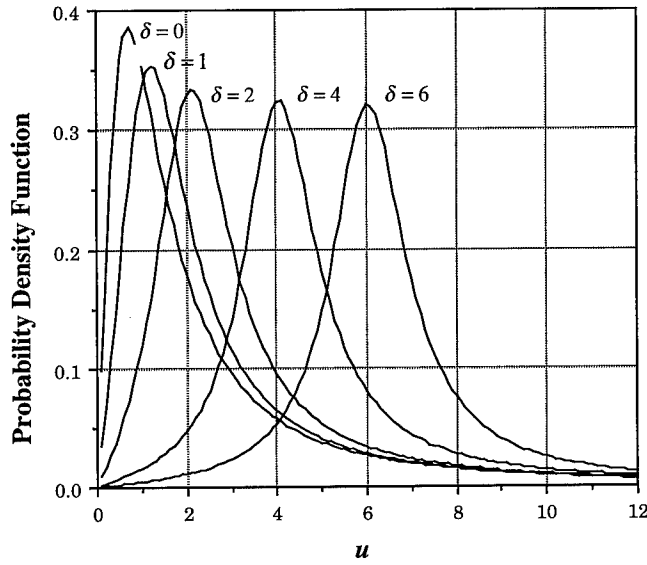


Fig. 8. Generalized Cauchy Envelope Distribution

To this point, the derivation of the density function is similar to that given by Tsihrintzis and Nikias<sup>7</sup>. For Gaussian noise, the corresponding density function is referred to as the Rician or generalized Rayleigh, so this density will be referred to as the generalized Cauchy envelope distribution. For a non-fading signal, the amplitude,  $A$ , is taken to be a constant over the observed samples.

Next the density function will be normalized by the negative first-order moment (NFM) of the noise

$$m_{-1} = 1/\gamma \quad (22)$$

Let  $u = z m_{-1}$  and  $\delta = A m_{-1}$ , and the density function becomes

$$f(u|\delta) = \frac{2u}{\pi(a-b)(a+b)^{1/2}} E\left(\sqrt{2b/(a+b)}\right) \quad (23)$$

where now  $a = u^2 + \delta^2 + 1$  and  $b = 2u\delta$ . This density is shown in Fig. 8 for different values of  $\delta$ .

The negative first-order moment of the generalized Cauchy distribution, or generalized Cauchy NFM is derived by integrating according to the following order

$$m_{-1}(A) = \int_{-\pi}^{\pi} \int_0^{\infty} \frac{1}{z} f(z, \phi | A, \theta) dz d\phi \quad (24)$$



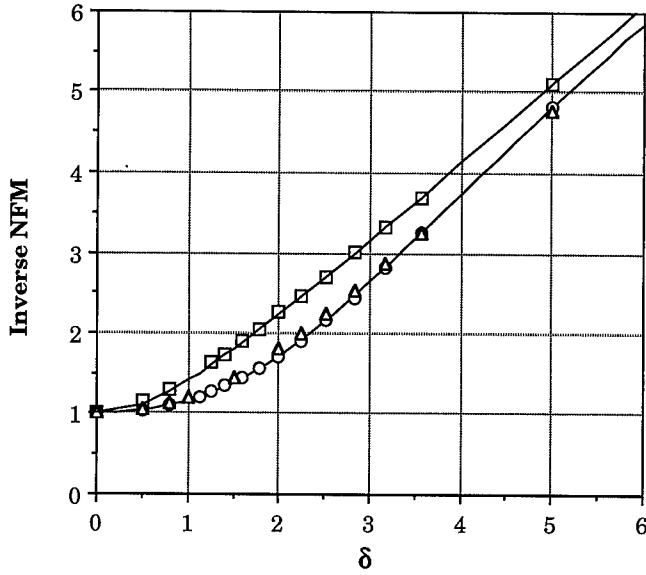


Fig. 9. Inverse Negative First-Order Moment for Generalized Cauchy (top) and Rician (bottom)

and then normalizing to get the closed-form solution

$$m_{-1}(\delta) = \frac{1}{\sqrt{\delta^2 + 1}} \quad (25)$$

The inverse of the NFM is shown in Fig. 3 for both the Generalized Cauchy density (top curve) and the Rician density (bottom curve). The NFM for the Rician<sup>1</sup> is also normalized by the noise-only NFM. The inverse NFM calculated using simulated Cauchy noise is given by squares; the inverse NFM using simulated Gaussian noise is given by circles. For each case, the calculated values closely match the estimated values. The triangles are inverse NFM estimates from H-Pol sea clutter with  $\alpha = 1.75$ . The sea clutter estimates closely match those from the Rician.

The probability density function of the phase is derived for a given sinewave amplitude and phase. The function is

$$f(\phi|A, \theta) = \int_0^\infty f(z, \phi|A, \theta) dz \quad (26)$$

where after manipulation and normalization, the Cauchy Phase PDF is

$$f(\phi|\delta, \theta) = \frac{1}{2\pi} \left( \frac{\sqrt{\delta^2 + 1} - \delta \cos(\phi - \theta)}{\delta^2 + 1 - \delta^2 \cos^2(\phi - \theta)} \right) \quad (27)$$

### 5. Statistical Error for Lower-Order Moments

The lower-order moments describe the parameters of the noise distribution. These moments are consistent, and the order of the moment can be chosen to reduce statistical or sampling error. The estimate of the  $p^{\text{th}}$  moment of an  $M^{\text{th}}$  variate, stable random process is given by

$$\hat{m}_p = \frac{1}{N} \sum_{n=1}^N |\vec{z}_n|^p \quad (28)$$

where

$$\vec{z}_n = (z_{1n}, z_{2n}, z_{3n}, \dots, z_{Mn}) \quad (29)$$

The estimator is unbiased, and for  $N$  independent samples, the normalized statistical error is

$$\frac{\sqrt{N \text{Var}[\hat{m}_p]}}{m_p} = \sqrt{\frac{m_{2p}}{m_p^2} - 1} \quad (30)$$

From Eq. (4), where  $-M < p < \alpha$

$$m_p = C(p, \alpha) \gamma^{\frac{p}{\alpha}} \quad (31)$$

and

$$C(p, \alpha) = 2^p \frac{\Gamma\left(\frac{p+M}{2}\right) \Gamma\left(1 - \frac{p}{\alpha}\right)}{\Gamma\left(\frac{M}{2}\right) \Gamma\left(1 - \frac{p}{2}\right)} \quad (32)$$

If alpha is known, then gamma can be estimated by

$$\hat{\gamma} = \left( \frac{\hat{m}_p}{C(p, \alpha)} \right)^{\frac{\alpha}{p}} \quad (33)$$

and the normalized sampling error for gamma is approximately

$$\frac{\sqrt{N \text{Var}[\hat{\gamma}]}}{\gamma} = \frac{\alpha}{|p|} \sqrt{\frac{m_{2p}}{m_p^2} - 1} \quad (34)$$

Fig. 10 shows this relationship for the bivariate case as alpha is varied from 1 to 2. As the noise becomes more impulsive, the negative lower-order moments have the least statistical error.

These relationships provide a way for estimating alpha and gamma. Since gamma is independent of moment order and alpha, a series of moments with different order can be calculated from the same set of samples. Gamma is estimated using Eq. (33) as alpha is varied. The alpha that gives a constant gamma estimate for the range of moment order is then the estimate of alpha. The moment order interval can be selected to minimize sampling error. The tails of the distribution, however, are not checked by this method.

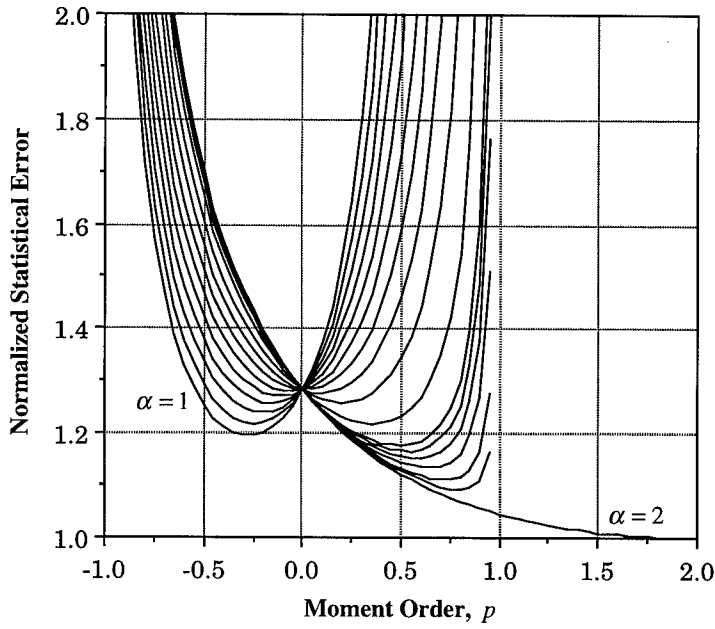


Fig. 10. Statistical Error for Gamma Estimates for Alpha = 1, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 1.95, 1.96, 1.97, 1.98, 1.99, 1.995, and 2

The statistics of short sea clutter segments can be examined using the moment method for estimating alpha and gamma. Fig. 11 and 12 show these results from 60 segments for H-pol and V-pol sea clutter with 2.1 m resolution in sea state 2. Each segment is 0.64 seconds long and consists of 400 samples (correlated) so each figure presents results from 38.4 seconds. The H-pol and V-pol data were collected on alternate samples, and the data are uncalibrated. The alpha estimated over the entire data set is 1.65 for the H-pol and 1.95 for the V-pol. Both data sets show significant variation in both alpha and gamma; the parameters are time varying. As expected, the H-pol data in Fig. 11 shows the most variation in both parameters, and the V-pol data in Fig. 12 has an alpha very close to 2.

## 6. Spectrum Analysis

For the non-Gaussian case when  $\alpha < 2$ , estimates of the auto-spectrum become inconsistent since the random process has infinite variance. Estimating the auto-spectrum using the Welch method<sup>8</sup> is the same as taking the second-order moment of the envelope

$$\hat{G}(\omega_j) = \frac{1}{N} \sum_{n=1}^N |X_n(\omega_j)|^2 \quad (35)$$

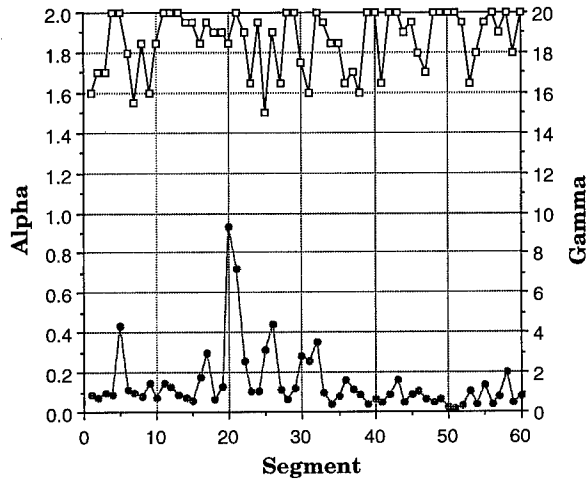


Fig. 11. Variation in Alpha and Gamma Over 0.64 Second Segments for Horizontally Polarized Sea Clutter

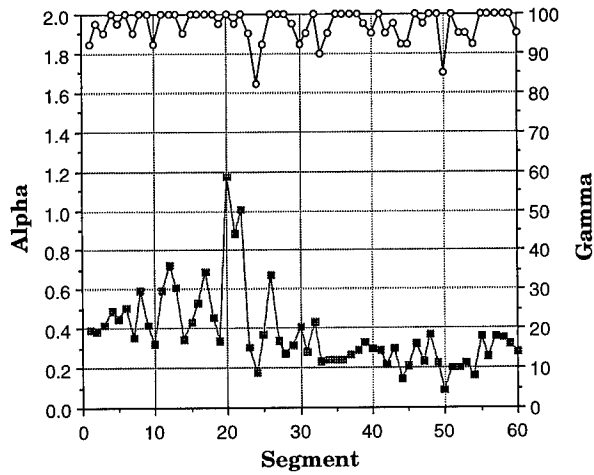


Fig. 12. Variation in Alpha and Gamma Over 0.64 Second Segments for Vertically Polarized Sea Clutter

where  $X_n(\omega_j)$  is the fast Fourier transform (FFT) of the  $n^{\text{th}}$  windowed, overlapping data segment.

$$X_n(\omega_j) = \frac{1}{K} \sum_{k=1}^K x((k-1 + \xi_n)\Delta t) \kappa(k\Delta t) e^{-i\omega_j \Delta t(k-1)} \quad (36)$$

where  $\omega_j = 2\pi(j-1)/K\Delta t$  and  $\xi_n = (n-1)S$ . Each segment is shifted  $S$  samples relative to the last segment,  $K$  is the FFT size,  $\kappa(k\Delta t)$  is the window function and  $\Delta t$  is the time between samples. To define a spectrum when  $\alpha < 2$ , the approach taken is to estimate alpha and gamma from the lower-order moments of  $|X_n(\omega_j)|$  as described in Sec. 5.

The time history,  $x(k\Delta t)$ , can either be a real-valued or a complex, stable random process. For either case, the FFT will produce bivariate isotropic, alpha-stable distributed samples for each data segment. For sea clutter, the data are complex.

An example using H-pol sea clutter to produce a Doppler spectrum is given in Fig. 13. The H-pol clutter was taken with 2.1 m down-range resolution in sea state 2 with a sample rate of 625 Hz. An artificial target was added to bin #16 (velocity index #16). The spectrum was produced from 373 segments, each 128 samples long with 50% overlap. The cosine data window was used; it was scaled to preserve the amplitude of the sinusoid

$$\frac{1}{K} \sum_{k=1}^K \kappa(k\Delta t) = 1 \quad (37)$$

Fig. 13 contains four overlaid plots: The conventional Doppler spectrum using Eq. (17) is plotted with the simulated target (open circles) and without the target (no symbols). The gamma Doppler spectrum Eq. (19), is plotted with the target (filled circles) and without the target (no symbols). The conventional and gamma spectra converge when  $\alpha = 2$ . Bin #16 without target is used as an example of inconsistent estimates in Fig. 1. The gamma spectrum in dB is defined as

$$\gamma(\omega_j)_{dB} = \frac{20}{\alpha(\omega_j)} \log \gamma(\omega_j) \quad (38)$$

The target is not seen in the conventional spectrum. Fig. 14 shows the alpha and gamma spectra without the target present. The alpha spectrum shows alpha's around 1.25 on the flanks of the gamma spectrum. The clutter Doppler spectra are clearly non-Gaussian. For bins less than -25 and greater than +25, the radar system noise with  $\alpha = 2$  dominates.

## 7. Conclusions

For high-resolution, horizontally polarized sea clutter, the alpha-stable distribution can give a much better fit to the envelope of the clutter returns than the generally accepted K-distribution. Sea clutter is used as an example. The negative first-order moment estimate of the envelope has various applications: estimate alpha for fitting stable distributions, normalize the envelope density function to fit distributions, estimate the median, normalize envelope detectors, detect targets, and set clipper level for detectors. The generalized Cauchy envelope distribution gives the distribution and the negative first-order moment for a sinewave plus Cauchy noise. The statistical error for the lower-order moments and estimates of gamma are developed. These lead to methods for estimating alpha and gamma that have application to short data segments. These methods are shown to produce gamma and alpha spectra that are consistent and give additional insight into the spectral fluctuations that occur with time.

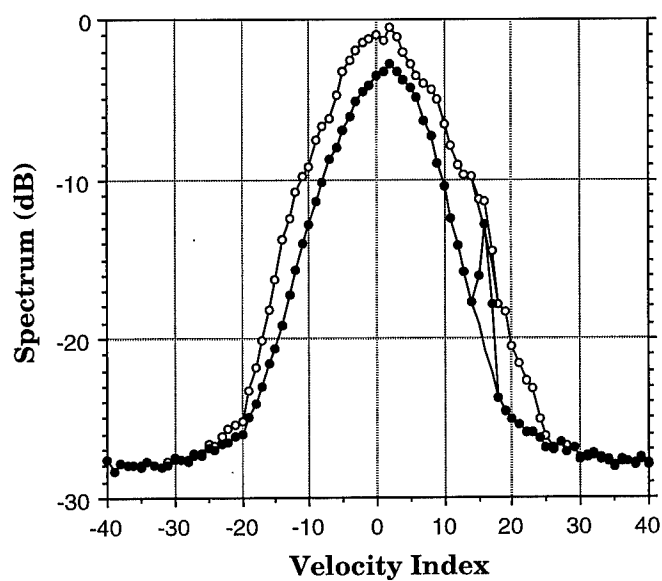


Fig. 13. Conventional and Gamma Doppler Spectrum for H-Pol Sea Clutter with and without Artificial Target in Bin 16

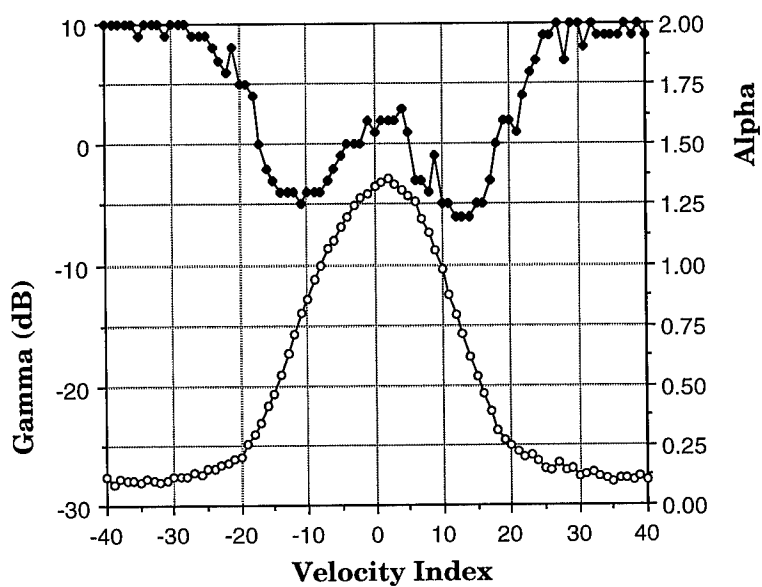


Fig. 14. Alpha and Gamma Doppler Spectrum for H-Pol Sea Clutter (no Target)

## 8. Administrative Information

This project was supported by the Carderock Division of the Naval Surface Warfare Center's In-house Laboratory Independent Research Program sponsored by the Office of Naval Research and administered by the Research Director, Code 0112 under Program Element 0601152N under NSWCCD Work Unit 1-7340-504.

## 9. References

1. A. D. Whalen, *Detection of Signals in Noise*, (Academic Press, New York, 1971).
2. M. Shao and C. L. Nikias, *Signal processing with fractional lower order moments: Stable processes and their applications*, (IEEE Proc., **81**, July 1993) pp. 986-1010.
3. C. L. Nikias and M. Shao, *Signal Processing with  $\alpha$ -Stable Distributions and Applications*, (John Wiley & Sons, New York, 1995).
4. G. Samorodnitsky and M. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*, (Chapman & Hall, New York, 1994).
5. T. Nohara and S. Haykin, *Canadian East Coast radar trials and the K-distribution*, (IEE Proc.-F, **138**, no. 2, April 1991) pp. 80-88.
6. B. Armstrong and H. Griffiths, *CFAR detection of fluctuating targets in spatially correlated K-distributed clutter*, (IEE Proc.-F, **138**, no. 2, April 1991) pp. 139-152.
7. G.A. Tsihrintzis and C.L. Nikias, *Incoherent receivers in alpha-stable impulsive noise*, (IEEE Trans. Signal Processing, **43**, no. 9, September 1995) pp. 2225-2229.
8. P. D. Welch, *The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms*, (IEEE Trans. Audio Electroacoust., **AU-15**, June 1967) pp. 70-73.

## DETRENDING TURBULENCE TIME SERIES WITH WAVELETS

EDGAR L ANDREAS

*U.S. Army Cold Regions Research and Engineering Laboratory  
72 Lyme Road  
Hanover, New Hampshire 03755-1290, U.S.A.  
E-Mail: eandreas@hanover-crrrel.army.mil*

and

GEORGE TREVIÑO

*Mechanical Engineering - Engineering Mechanics Department  
Michigan Technological University  
Houghton, Michigan 49931, U.S.A.  
E-Mail: gtrevino@mtu.edu*

### ABSTRACT

Wavelets are a new class of basis functions that are finding wide use for analyzing and interpreting turbulence data. Here we describe a new use for wavelets: identifying trends in turbulence time series. The general turbulence signal we consider has a quadratic trend. We use the inverted Haar wavelet and the elephant wavelet, respectively, to estimate the first-order and second-order coefficients in the trend polynomial. Unlike usual wavelet applications, however, we use only one dilation scale,  $L$ , where  $L$  is the total length of the turbulence series. Our analysis shows that wavelet trend detection is roughly half as accurate as least-squares trend detection when accuracy is evaluated in terms of the mean-square error in estimates of the first-order and second-order trend coefficients. But wavelet detection is more than twice as efficient as least-squares detection in the sense that it requires fewer than half the number of floating point operations of least-squares regression to yield the three coefficients of the quadratic trend polynomial. We demonstrate wavelet trend detection first with artificial data and then with various data collected in the atmospheric surface layer. Lastly, we provide guidelines on when linear and quadratic trends are "significant" enough to require removal from turbulence series.

### 1. Introduction

Removing the trend from a turbulence series is an essential but arcane art. Bendat and Piersol (1971, p. 291) and Panofsky and Dutton (1984, p. 175), for example, explain that failure to remove a trend may result in distorted spectra or correlation functions, especially at low frequencies or longer time lags. Although many methods have been used for trend removal (e.g., Bendat and Piersol 1971, 288 ff.; Priestley 1981, p. 587 ff.; Panofsky and Dutton 1984, p. 87 ff.), none has emerged as the standard method, at least in the atmo-



spheric turbulence community (Panofsky and Dutton 1984, p. 175), and particularly for nonstationary turbulence.

To backtrack for a moment, our working definition here is that a trend is any component of a signal with a period longer than the length of the record. In essence, a series with a trend is nonstationary (Jenkins and Watts 1968, p. 151) and thus an appropriate subject for this volume. Although we focus here on turbulence time series, everything we say applies directly to spatial turbulence series, collected by aircraft, for example. Simply replace "time" with "distance," "frequency" with "wavenumber," and "period" with "wavelength."

One of the most common methods for detecting trends in time series is least-squares regression. We offer here, however, an alternative to least-squares regression based on wavelets. Wavelets are a new family of basis functions (Daubechies 1988, Strang 1989, Farge 1992) that are becoming increasingly popular for a wide range of meteorological applications (e.g., Gao and Li 1993, Collineau and Brunet 1993, Mahrt and Gibson 1992, Mahrt and Howell 1994, Katul and Parlange 1994, Turner and Leclerc 1994). We show how simple it is to use wavelets to detect trends in turbulence time series. As part of this demonstration, we develop the mathematics for using wavelets to detect first-order and second-order trends, prove that wavelets yield unbiased and consistent estimators of the trend, offer guidelines on when it is necessary to remove trends, and present some examples of trend detection using wavelets.

Table 1. Consider two scenarios: a turbulence time series  $f(t)$  with a linear trend, and a turbulence time series  $g(t)$  with a quadratic trend. If  $N$  is the total number of points in the time series, the entries show how many floating-point operations are necessary to estimate the trend polynomial (the  $\mu$ 's) with our wavelet detection scheme and with least-squares regression. The wavelet operation counts come from the discrete forms of the expressions for  $\hat{\mu}_0$ ,  $\hat{\mu}_1$ , and  $\hat{\mu}_2$  developed in Section 3. The least-squares operation counts come simply from counting operations in Bevington's (1969, p. 99 ff. and p. 137) expressions for  $\hat{\mu}_{0,LS}$ ,  $\hat{\mu}_{1,LS}$ , and  $\hat{\mu}_{2,LS}$ , after we first recognize that sums of the first, second, third, and fourth powers of  $t_i = \Delta i$  are summable series requiring far fewer than  $N$  operations.

<b>Linear: <math>\tilde{f}(t) = f(t) + \mu_0 + \mu_1 t</math></b>	
Wavelets—	$2N + 8$
Least-Squares Regression—	$3N + 14$
<b>Quadratic: <math>\tilde{g}(t) = g(t) + \mu_0 + \mu_1 t + \mu_2 t^2</math></b>	
Wavelets—	$3N + 23$
Least-Squares Regression—	$6N + 73$

Since least-squares regression is the primary alternative to our proposed wavelet trend-detection scheme, we also examine the expected accuracy of both least-squares regression and wavelet detection. Least-squares regression proves to be better than wavelet detection in terms of the mean-square error in the estimated coefficients of the polynomial trend, but wavelet detection has a computational advantage. Table 1 gives the bottom-line benefits of wavelet detection. If  $N$  is the number of points in a turbulence time series, wavelet detection requires roughly  $N$  fewer floating-point operations than least-squares regression for detecting a linear trend. In testing for a quadratic trend, wavelet detection requires less than half as many operations and is, therefore, more than twice as fast as least-squares regression.

## 2. Mathematical Foundation

Because we want to detect trends up to second order, we use two wavelets in our trend-detection scheme, the inverted Haar wavelet (Figure 1) and the elephant wavelet (Figure 2). The inverted Haar wavelet at time  $t$ ,  $I(t, L)$ , is

$$A^{-1}I(t, L) = -1 \quad \text{if } -L/2 \leq t < 0, \quad (2.1a)$$

$$= 1 \quad \text{if } 0 \leq t \leq L/2, \quad (2.1b)$$

$$= 0 \quad \text{elsewhere.} \quad (2.1c)$$

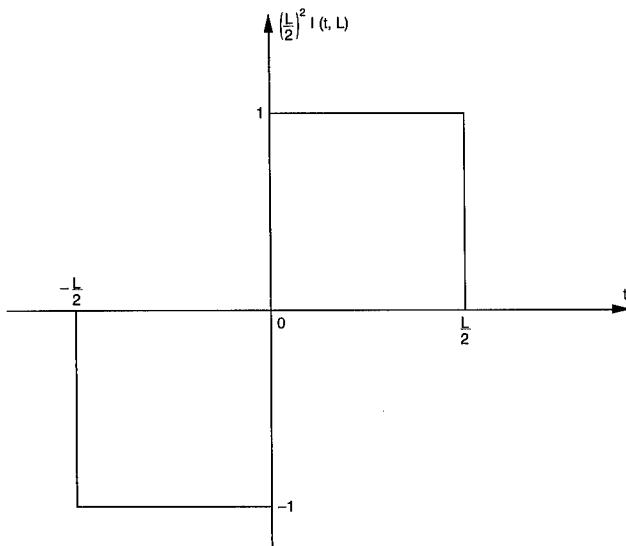


Fig. 1. Inverted Haar wavelet.

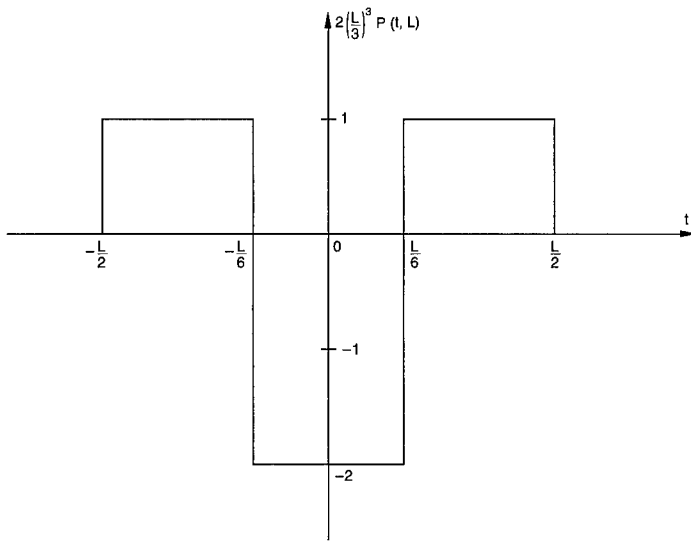


Fig. 2 Elephant wavelet.

Here

$$A = \left(\frac{2}{L}\right)^2, \quad (2.2)$$

where  $L$  is the dilation scale of the wavelet—the length of time it encompasses. The elephant wavelet,  $P(t, L)$ , is

$$B^{-1}P(t, L) = 1 \quad \text{if } -L/2 \leq t < -L/6, \quad (2.3a)$$

$$= -2 \quad \text{if } -L/6 \leq t < L/6, \quad (2.3b)$$

$$= 1 \quad \text{if } L/6 \leq t \leq L/2, \quad (2.3c)$$

$$= 0 \quad \text{elsewhere,} \quad (2.3d)$$

where

$$B = \frac{1}{2} \left(\frac{3}{L}\right)^3. \quad (2.4)$$

Although this is just an inverted French hat wavelet, we prefer the less cumbersome appellation elephant wavelet because the wavelet looks like an elephant to us: it has a long trunk and two big ears.

Admittedly, we are not being strictly rigorous in referring to Eq. (2.1) and (2.3) as wavelets. First, we use only one dilation scale in our trend-detection scheme,  $L$ , where  $L = \Delta N$  is the total length of the time series and  $\Delta$  is the sampling interval. Second, we ignore energy conservation considerations, which would dictate other values for  $A$  and  $B$  (Farge 1992). But, by Daubechies's (1988) definition (also Gamage and Hagelberg 1993), to be termed a wavelet, a function need only have compact support (be nonzero only over a finite domain) and satisfy an admissibility condition (have zero mean), which Eq. (2.1) and (2.3) do.

Although we first deal with Eq. (2.1) and (2.3) as continuous functions, for practical time series analysis we must use the discrete forms. For the inverted Haar wavelet, the discrete form is

$$\left(\frac{\Delta N}{2}\right)^2 I(i, N) = -1 \quad \text{for } 1 \leq i \leq N/2, \quad (2.5a)$$

$$= 1 \quad \text{for } N/2 + 1 \leq i \leq N, \quad (2.5b)$$

$$= 0 \quad \text{otherwise.} \quad (2.5c)$$

For the elephant wavelet, it is

$$2 \left(\frac{\Delta N}{3}\right)^3 P(i, N) = 1 \quad \text{for } 1 \leq i \leq N/3, \quad (2.6a)$$

$$= -2 \quad \text{for } N/3 + 1 \leq i \leq 2N/3, \quad (2.6b)$$

$$= 1 \quad \text{for } 2N/3 + 1 \leq i \leq N, \quad (2.6c)$$

$$= 0 \quad \text{otherwise.} \quad (2.6d)$$

Because in our wavelet-based detrending scheme we ultimately apply both Eq. (2.5) and (2.6) to the same time series,  $N$  must be a multiple of six to facilitate segmenting the series.

### 3. Accuracy Considerations

#### 3.1. Unbiased Estimators

Because the advantage that wavelet trend detection has over least-squares regression is in its efficiency in detecting quadratic trends (see Table 1), the general turbulence signal we consider is

$$\tilde{g}(t) = g(t) + \mu_0 + \mu_1 t + \mu_2 t^2. \quad (3.1)$$

Here,  $t$  is time;  $\tilde{g}$  is the instantaneous (measured) value of the time series;  $g$  is the (zero-mean) turbulence component; and  $\mu_0$ ,  $\mu_1$ , and  $\mu_2$  are the coefficients of the quadratic trend polynomial. The  $\tilde{g}$  series is defined (measured) only for  $0 \leq t \leq L$ . The purpose of trend detection is to estimate  $\mu_0$ ,  $\mu_1$ , and  $\mu_2$  and then to isolate the turbulence signal  $g(t)$ . All that follows could also be adapted to a series with a linear trend by setting  $\mu_2 = 0$ .

In general, convolving a wavelet with a time series defines a wavelet coefficient (e.g., Farge 1992). Specifically, convolving  $\tilde{g}(t)$  with the elephant wavelet yields the wavelet coefficient  $\chi$ ,

$$\chi(t, L) = \int_{-\infty}^{\infty} \tilde{g}(s) P(s - t, L) ds. \quad (3.2)$$

Using Eq. (2.3) gives

$$\chi(L/2, L) = B \int_0^{L/3} \tilde{g}(s) ds - 2B \int_{L/3}^{2L/3} \tilde{g}(s) ds + B \int_{2L/3}^L \tilde{g}(s) ds \quad (3.3)$$

since  $P(t, L)$  is zero outside  $[0, L]$ . Notice, in most wavelet analyses, the support of the wavelet  $L$  is much shorter than the length of the time series. Thus, moving the wavelet through the time series according to Eq. (3.2) would yield a plethora of wavelet coefficients indexed to  $t$  and  $L$ . But in our analysis, Eq. (3.3) yields only one  $\chi$  value, the one at time  $L/2$ , since the length of  $P$  exactly matches the length of  $\tilde{g}$ .

With two changes of variables, Eq. (3.3) becomes

$$\chi(L/2, L) = B \int_0^{L/3} ds \left[ \tilde{g}(s) - 2\tilde{g}\left(s + \frac{L}{3}\right) + \tilde{g}\left(s + \frac{2L}{3}\right) \right]. \quad (3.4)$$

Substitute Eq. (3.1) in for  $\tilde{g}$  here. The result is

$$\chi(L/2, L) = B \int_0^{L/3} ds \left[ g(s) - 2g\left(s + \frac{L}{3}\right) + g\left(s + \frac{2L}{3}\right) \right] + \mu_2. \quad (3.5)$$

Since  $g(t)$  has zero-mean, the integral in Eq. (3.5) should be near zero. Thus,  $\chi(L/2, L)$  is an estimator of  $\mu_2$ ; call it  $\hat{\mu}_2$ . Formally, we can take the expectation of  $\hat{\mu}_2$ ,  $E[\hat{\mu}_2]$ , to confirm this;

$$E[\hat{\mu}_2] = BE \left[ \int_0^{L/3} ds \left[ g(s) - 2g\left(s + \frac{L}{3}\right) + g\left(s + \frac{2L}{3}\right) \right] \right] + E[\mu_2]. \quad (3.6)$$

Since  $E[g(t)] = 0$  by definition,

$$E[\hat{\mu}_2] = \mu_2. \quad (3.7)$$

That is,  $\hat{\mu}_2 = \chi(L/2, L)$  is an unbiased estimator of  $\mu_2$ .

Notice, the integrand in Eq. (3.4) resembles the second-difference operator (Haltiner and Williams 1980, p. 109; Press et al. 1994, p. 161). Therefore, it is not surprising that operating on a time series with the elephant wavelet yields a coefficient that estimates the second-order trend in the time series.

Next, convolve  $\tilde{g}(t)$  with the inverted Haar wavelet. This yields the wavelet coefficient  $X$ ,

$$X(t, L) = \int_{-\infty}^{\infty} \tilde{g}(s) I(s-t, L) ds. \quad (3.8)$$

Using Eq. (2.1) and transforming variables gives

$$X(L/2, L) = A \int_0^{L/2} ds \left[ \tilde{g}\left(s + \frac{L}{2}\right) - \tilde{g}(s) \right]. \quad (3.9)$$

Substituting Eq. (3.1) for  $\tilde{g}$  and doing some algebra and two integrations yields

$$X(L/2, L) = A \int_0^{L/2} ds \left[ g\left(s + \frac{L}{2}\right) - g(s) \right] + \mu_1 + \mu_2 L. \quad (3.10)$$

Consequently, since the integral should be near zero,  $X(L/2, L)$  combined with Eq. (3.4) provides an estimate of  $\mu_1$ —call it  $\hat{\mu}_1$ —where

$$\hat{\mu}_1 = X(L/2, L) - \hat{\mu}_2 L. \quad (3.11)$$

Or, substituting Eq. (3.10) in (3.11),

$$\hat{\mu}_1 = \mu_1 + A \int_0^{L/2} ds \left[ g\left(s + \frac{L}{2}\right) - g(s) \right] - (\hat{\mu}_2 - \mu_2)L. \quad (3.12)$$

To formally establish the validity of this estimator, take the expected value of Eq. (3.12);

$$E[\hat{\mu}_1] = E[\mu_1] + A \int_0^{L/2} ds E \left[ g\left(s + \frac{L}{2}\right) - g(s) \right] - L E[\hat{\mu}_2 - \mu_2], \quad (3.13)$$

or

$$E[\hat{\mu}_1] = \mu_1, \quad (3.14)$$

since  $g$  has zero mean and because  $\hat{\mu}_2$  is an unbiased estimator of  $\mu_2$ . Consequently,  $\hat{\mu}_1$  is an unbiased estimator of  $\mu_1$ .

Notice that the integrand in Eq. (3.9) resembles the forward-difference or first-difference operator (Haltiner and Williams 1980, p. 109; Press et al. 1994, p. 161). Thus, it is clear why operating on a series with the inverted Haar wavelet yields a quantity related to the coefficient of the linear trend in the time series.

In essence, Eq. (3.9) shows that we can estimate the linear trend in a time series by simply subtracting the average of the first half of the series from the average of the second half. Bendat and Piersol (1971, p. 289) and Treviño (1982) present similar methods for estimating the slope of a linear trend, although they use only portions of the time series, not the entire time series as we do.

The sample average of  $\tilde{g}$  is

$$\bar{\tilde{g}} = \frac{1}{L} \int_0^L \tilde{g}(s) ds. \quad (3.15)$$

Substituting Eq. (3.1) for  $\tilde{g}$ , we get

$$\bar{\tilde{g}} = \frac{1}{L} \int_0^L g(s) ds + \mu_0 + \frac{1}{2} \mu_1 L + \frac{1}{3} \mu_2 L^2. \quad (3.16)$$

Again, because  $g$  should experimentally have zero mean, an estimator for  $\mu_0$  is

$$\hat{\mu}_0 = \bar{\tilde{g}} - \frac{1}{2} \hat{\mu}_1 L - \frac{1}{3} \hat{\mu}_2 L^2. \quad (3.17)$$

Or, from Eq. (3.16),

$$\hat{\mu}_0 = \mu_0 + \frac{1}{L} \int_0^L g(s) ds - \frac{1}{2} L (\hat{\mu}_1 - \mu_1) - \frac{1}{3} L^2 (\hat{\mu}_2 - \mu_2). \quad (3.18)$$

On taking the expected value of this, we see

$$E[\hat{\mu}_0] = E[\mu_0] + \frac{1}{L} \int_0^L E[g(s)] ds - \frac{1}{2} L E[\hat{\mu}_1 - \mu_1] - \frac{1}{3} L^2 E[\hat{\mu}_2 - \mu_2]. \quad (3.19)$$

Thus,

$$E[\hat{\mu}_0] = \mu_0, \quad (3.20)$$

since  $g$  has zero mean and  $\hat{\mu}_1$  and  $\hat{\mu}_2$  are unbiased estimators of  $\mu_1$  and  $\mu_2$ , respectively. That is, our wavelet-derived estimate of  $\mu_0$  is also an unbiased estimator.

### 3.2. Consistent Estimators

A statistic for evaluating the accuracy of the estimators we defined in the last section is the mean-square error (Treviño 1982) or error variance (Lenschow et al. 1994). With  $\hat{\mu}_2$  as an example, the mean-square-error (mse) operator is defined as

$$\text{MSE}[\hat{\mu}_2] \equiv E[(\hat{\mu}_2 - \mu_2)^2]. \quad (3.21)$$

In Appendix A, we derive the mse of each estimator that we defined in the last section. These are all functions of the integral time scale  $\mathfrak{S}$ , which is defined as

$$\mathfrak{S} = \sigma_g^{-2} \int_0^\infty C_g(\tau) d\tau. \quad (3.22)$$

Here,  $\sigma_g^2$  is the population variance of the  $g$  time series, and  $C_g$  is the autocorrelation of  $g$ ,

$$C_g(\tau) = E[g(t)g(t + \tau)], \quad (3.23)$$

where  $\tau$  is the time lag. We also recognize  $C_g(\tau)/\sigma_g^2$  as the normalized autocorrelation function.

In Appendix A we show that

$$\text{MSE}[\hat{\mu}_0] = 21.25 \left( \frac{\mathfrak{S}}{L} \right) \sigma_g^2, \quad (3.24a)$$

$$\text{MSE}[\hat{\mu}_1] = \frac{653}{L^2} \left( \frac{\mathfrak{S}}{L} \right) \sigma_g^2, \quad (3.24b)$$

$$\text{MSE}[\hat{\mu}_2] = \frac{729}{L^4} \left( \frac{\mathfrak{S}}{L} \right) \sigma_g^2. \quad (3.24c)$$

If  $\mathfrak{S}$  is finite and much smaller than the total series length  $L$ ,  $\hat{\mu}_0$ ,  $\hat{\mu}_1$ , and  $\hat{\mu}_2$  are all consistent estimators: The mse of each goes to zero as the series length (number of samples) increases.

Because, from our point of view, least-squares regression is the primary alternative to our wavelet-based trend detection, in Appendix B we likewise investigate the statistical properties of the least-squares estimators of  $\mu_0$ ,  $\mu_1$ , and  $\mu_2$ . Call these estimators  $\hat{\mu}_{0,LS}$ ,  $\hat{\mu}_{1,LS}$ , and  $\hat{\mu}_{2,LS}$ . In Appendix B, we show that—as with  $\hat{\mu}_0$ ,  $\hat{\mu}_1$ , and  $\hat{\mu}_2$ — $\hat{\mu}_{0,LS}$ ,  $\hat{\mu}_{1,LS}$ , and  $\hat{\mu}_{2,LS}$  are unbiased estimators of  $\mu_0$ ,  $\mu_1$ , and  $\mu_2$ , respectively. We also find that

$$\text{MSE}[\hat{\mu}_{0,LS}] = 18 \left( \frac{\mathfrak{S}}{L} \right) \sigma_g^2, \quad (3.25a)$$

$$\text{MSE}[\hat{\mu}_{1,LS}] = \frac{384}{L^2} \left( \frac{\mathfrak{S}}{L} \right) \sigma_g^2, \quad (3.25b)$$

$$\text{MSE}[\hat{\mu}_{2,LS}] = \frac{360}{L^4} \left( \frac{\mathfrak{S}}{L} \right) \sigma_g^2. \quad (3.25c)$$



That is, the least-squares estimators all provide consistent estimates of the quantities they are intended to predict. On comparing Eq. (3.24) and (3.25), we see that  $\text{MSE}[\hat{\mu}_{0,LS}]$  is somewhat less than  $\text{MSE}[\hat{\mu}_0]$ . The mse's of the other two least-squares estimators are roughly half as large as the mse's of the corresponding wavelet-based estimators. Thus, in the mse sense, least-squares trend detection is better than wavelet-based trend detection. But remember Table 1; this factor of two increase in accuracy comes at the expense of a factor of two increase in computing operations.

#### 4. Tests with Artificial Data

To test our wavelet-based trend detection scheme and to verify the statistical properties of the estimators that we derived in the last section, we constructed some artificial time series with known quadratic trends. Each time series is of the form Eq. (3.1), where we have specified  $\mu_0$ ,  $\mu_1$ , and  $\mu_2$ . We generated  $g(t)$  by randomly sampling from a Gaussian probability distribution that had zero mean and specified standard deviation  $\sigma_g$ .

For each set of  $\mu_0$ ,  $\mu_1$ , and  $\mu_2$ , we created 40 artificial time series. The first 20 series each consisted of 3000  $\tilde{g}$  values (this is  $N$ ) but differed because the sequence of random  $g$  values differed. The second 20 time series consisted of 12,000  $\tilde{g}$  values so we could see how quadrupling  $L$  affects  $\hat{\mu}_0$ ,  $\hat{\mu}_1$ , and  $\hat{\mu}_2$  [see Eq. (3.24)].

Figure 3 shows a typical 12,000-point time series with our first quadratic trend. A typical 3000-point time series with the same trend would just be the left one-quarter of this figure. The units are unimportant here since  $\tilde{g}$  could really be any quantity and  $t$  in Eq. (3.1) could be replaced with  $i$ , the index of the time step.

Figure 4 shows our wavelet-based estimates of  $\hat{\mu}_0$ ,  $\hat{\mu}_1$ , and  $\hat{\mu}_2$  for each run in the twenty 3000-point series and in the twenty 12,000-point series. After we computed  $\hat{\mu}_0$ ,  $\hat{\mu}_1$ , and  $\hat{\mu}_2$  for each artificial series, we detrended  $\tilde{g}$  and then computed the sample standard deviation,  $s$ , of the remaining  $g$  series. These  $s$  values also appear in Figure 4.

The variability of the estimates in Figure 4, especially in the 3000-point panels, shows how the sequence of data can lead to random scatter. The estimates are, however, evenly scattered both above and below the value being estimated. Thus, our analysis of these artificial data supports our prediction that  $\hat{\mu}_0$ ,  $\hat{\mu}_1$ , and  $\hat{\mu}_2$  are unbiased estimators.

The startling differences between the 3000-point and 12,000-point panels in Figure 4 is the record-length effect. We showed in Eq. (3.24) that  $\hat{\mu}_0$ ,  $\hat{\mu}_1$ , and  $\hat{\mu}_2$  should be consistent estimators. The differences between the left and right panels confirms this;  $\hat{\mu}_0$ ,  $\hat{\mu}_1$ , and  $\hat{\mu}_2$  all are much closer to the true values of  $\mu_0$ ,  $\mu_1$ , and  $\mu_2$  when the record length is four times longer. Even  $s$  is a better estimate of the true population standard deviation with the longer record because the better estimates of  $\mu_0$ ,  $\mu_1$ , and  $\mu_2$  facilitate more accurate trend removal and, consequently, a better estimate of the terms in the  $g$  series.

Figures 5 and 6 repeat this story using artificial data with stronger first-order and second-order trends. Again, the left one-quarter of Figure 5 is typical of the time series used

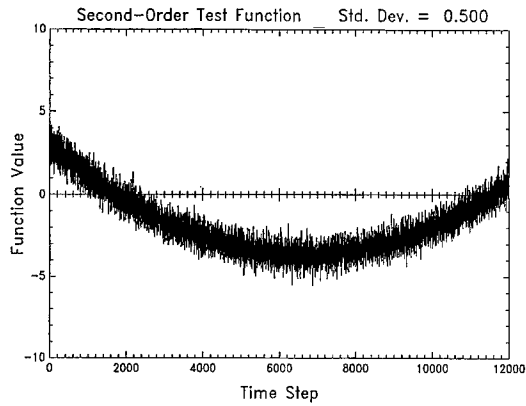


Fig. 3. Typical artificial time series generated with  $\sigma_g = 0.500$ ,  $\mu_0 = 3.00$ ,  $\mu_1 = -2.00 \times 10^{-3}$ , and  $\mu_2 = 1.50 \times 10^{-7}$ .

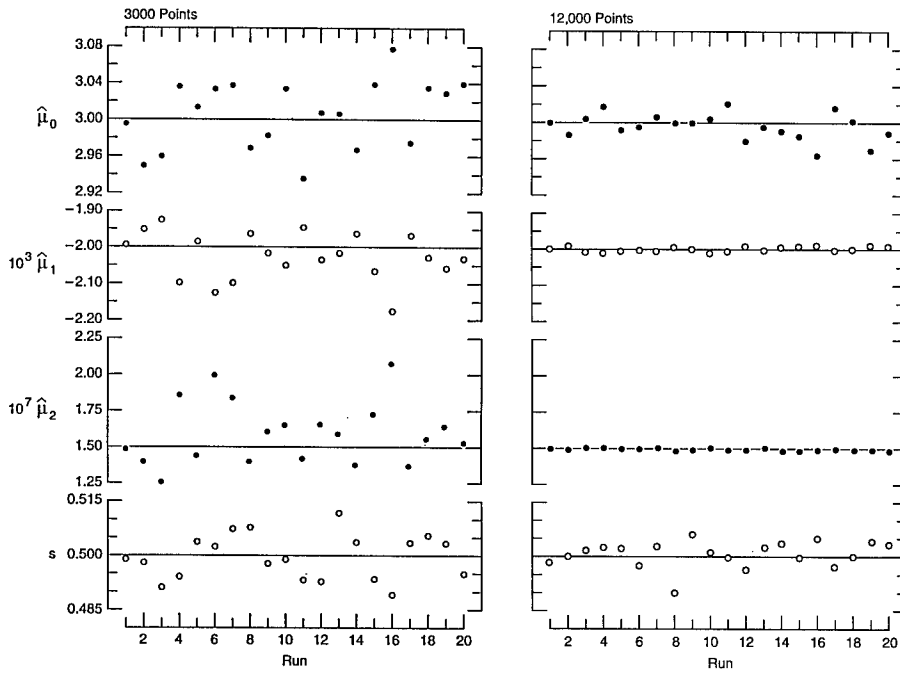


Fig. 4. Wavelet-based estimators  $\hat{\mu}_0$ ,  $\hat{\mu}_1$ , and  $\hat{\mu}_2$  for twenty 3000-point series and for twenty 12,000-point random series, each with the quadratic trend depicted in Fig. 3. In the lowest panel,  $s$  is the sample standard deviation of the turbulence signal ( $g$ ) computed using  $\hat{\mu}_0$ ,  $\hat{\mu}_1$ , and  $\hat{\mu}_2$  to remove the trend from the  $\tilde{g}$  series. In each panel, the horizontal line is the prescribed value of the quantity being estimated.

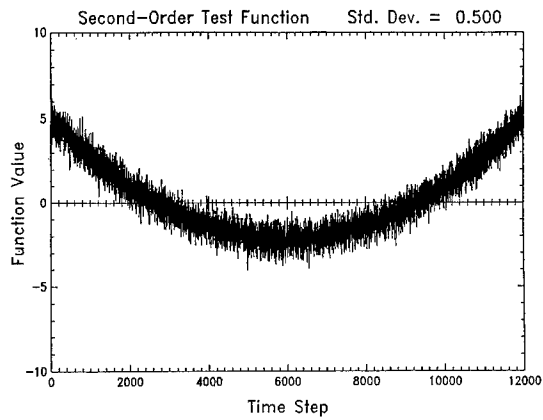


Fig. 5. Typical artificial time series generated with  $\sigma_g = 0.500$ ,  $\mu_0 = 5.00$ ,  $\mu_1 = -2.40 \times 10^{-3}$ , and  $\mu_2 = 2.00 \times 10^{-7}$ .

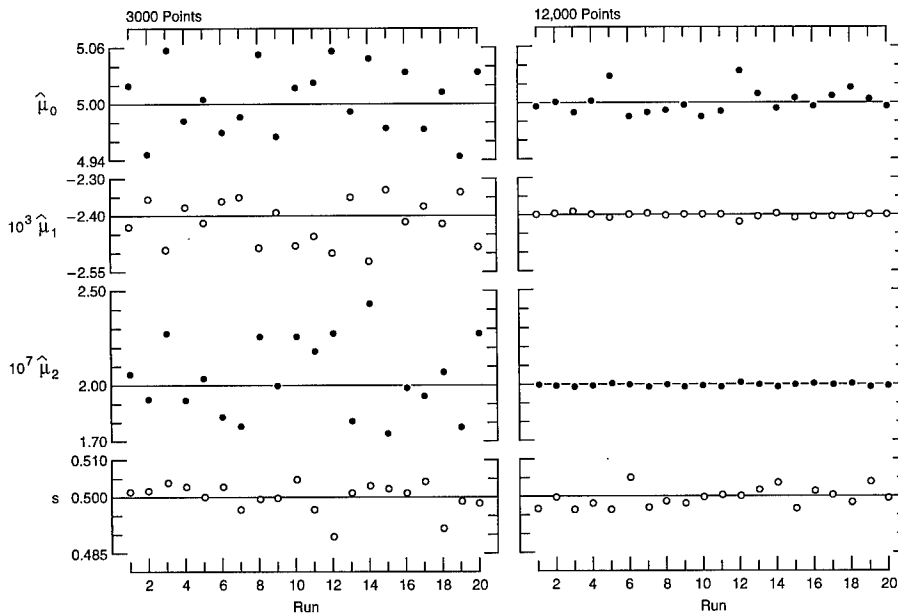


Fig. 6. As in Fig. 4, except this is the wavelet-based analysis of the time series in Fig. 5.

for the 3000-point analyses depicted in Figure 6. By eye it is difficult to see a second-order trend in this portion of the time series; but the wavelet analysis, nevertheless, can still estimate  $\mu_2$  to within about  $\pm 10\%$  from these 3000 points (see  $\hat{\mu}_2$  panel on left side of Figure 6).

In comparing the left and right panels of Figure 6, we see results as in Figure 4.  $\hat{\mu}_0$ ,  $\hat{\mu}_1$ , and  $\hat{\mu}_2$  are unbiased estimators of  $\mu_0$ ,  $\mu_1$ , and  $\mu_2$ , respectively. And all are consistent estimators: with the record length increased by a factor of four, all are much nearer to the values they are intended to estimate. The sample standard deviation,  $s$ , also estimates  $\sigma_g$  better for the longer time series because the better estimates of  $\hat{\mu}_0$ ,  $\hat{\mu}_1$ , and  $\hat{\mu}_2$  provide more accurate trend removal and thus a better determination of the  $g$  series.

## 5. Criteria for Removing Trends

As we mentioned, many sources discuss techniques for trend detection, and most explain the necessity for removing "significant" trends. Most also, however, fail to define "significant." We feel that the trend-detection problem really has two parts: (1) estimate the coefficients of the polynomial trend; (2) decide whether to remove this trend from the time series. Without addressing part 2, a trend-detection scheme is incomplete. Hence, here we provide some quantitative criteria for deciding when to remove an estimated trend.

Consider the discrete version of the time series in Eq. (3.1),

$$\tilde{g}_i = g_i + \mu_0 + \mu_1 i\Delta + \mu_2 (i\Delta)^2, \quad (5.1)$$

where  $i$  is the sample index,  $1 \leq i \leq N$ , and  $\Delta$  is the sampling interval. For example, with a sonic anemometer/thermometer operating in the atmospheric surface layer,  $\Delta$  might be 0.1 second.

By analogy with Eq. (3.15), the sample average of  $\tilde{g}$  in discrete form is

$$\bar{\tilde{g}} = \frac{1}{N} \sum_{i=1}^N \tilde{g}_i, \quad (5.2)$$

or

$$\bar{\tilde{g}} = \frac{1}{N} \sum_{i=1}^N [g_i + \mu_0 + \mu_1 i\Delta + \mu_2 (i\Delta)^2]. \quad (5.3)$$

We can easily sum these  $i$  series to get

$$\bar{\tilde{g}} = \frac{1}{N} \sum_{i=1}^N g_i + \mu_0 + \frac{\mu_1 \Delta}{2} (N+1) + \frac{\mu_2 \Delta^2}{6} (N+1)(2N+1), \quad (5.4)$$

which is analogous to Eq. (3.16).

By definition, the unbiased sample variance of  $\tilde{g}$  is

$$s_{\tilde{g}}^2 = \frac{1}{N-1} \sum_{i=1}^N (\tilde{g}_i - \bar{\tilde{g}})^2. \quad (5.5)$$

From Eq. (5.1) and (5.4), this is

$$s_{\tilde{g}}^2 = \frac{1}{N-1} \sum_{i=1}^N \left[ \left( g_i - \frac{1}{N} \sum_{k=1}^N g_k \right) + \mu_1 \Delta \left( i - \frac{N+1}{2} \right) + \mu_2 \Delta^2 \left( i^2 - \frac{(N+1)(2N+1)}{6} \right) \right]^2. \quad (5.6)$$

On multiplying this out and taking the expected value, we get

$$\begin{aligned} \sigma_{\tilde{g}}^2 = \sigma_g^2 &+ \frac{1}{N-1} \sum_{i=1}^N \left\{ \mu_1^2 \Delta^2 \left[ i^2 - i(N+1) + \frac{1}{4}(N+1)^2 \right] \right. \\ &+ 2\mu_1 \mu_2 \Delta^3 \left[ i^3 - \frac{i^2}{2}(N+1) - \frac{i}{6}(N+1)(2N+1) + \frac{1}{12}(N+1)^2(2N+1) \right] \\ &\left. + \mu_2^2 \Delta^4 \left[ i^4 - \frac{i^2}{3}(N+1)(2N+1) + \frac{1}{36}(N+1)^2(2N+1)^2 \right] \right\}. \quad (5.7) \end{aligned}$$

Summing the  $i$  series finally yields

$$\sigma_{\tilde{g}}^2 = \sigma_g^2 + \frac{\Delta^2 N(N+1)}{180} [15\mu_1^2 + 30\mu_1 \mu_2 \Delta(N+1) + \mu_2^2 \Delta^2 (2N+1)(8N+11)], \quad (5.8)$$

where  $\sigma_{\tilde{g}}^2$  and  $\sigma_g^2$  are, respectively, the population variances of the  $\tilde{g}$  and  $g$  series. Equation (5.8) establishes how badly a series with a trend misrepresents the variance of the actual turbulence signal.

Suppose the instrument we are using to sample  $g$  has a resolution  $r$ . For example, sonic anemometer/themometers made by Applied Technologies Inc. (ATI; Boulder, Colorado) provide a digital output reported to the nearest cm/s for components of the wind vector and to the nearest  $0.01^\circ\text{C}$  for temperature. For such instruments,  $r$  would then presumably be 1 cm/s for wind speed measurements and  $0.01^\circ\text{C}$  for temperature measurements. Since we cannot measure  $\sigma_g^2$  with more precision than our instrument provides, we cannot confidently identify a trend in our data unless

$$|\sigma_{\tilde{g}}^2 - \sigma_g^2| > r^2. \quad (5.9)$$

From Eq. (5.8) this criterion requires that we detrend our data only when

$$\left| \frac{\Delta^2 N(N+1)}{180} [15\hat{\mu}_1^2 + 30\hat{\mu}_1\hat{\mu}_2\Delta(N+1) + \hat{\mu}_2^2\Delta^2(2N+1)(8N+11)] \right| \equiv D_2^2 > r^2. \quad (5.10)$$

If we suspect that our time series has only a linear trend, not a quadratic one, the criterion for detrending is [from Eq. (5.8) with  $\mu_2 = 0$ ]

$$\left| \frac{\hat{\mu}_1^2 \Delta^2 N(N+1)}{12} \right| \equiv D_1^2 > r^2. \quad (5.11)$$

In summary—and this is valid for any trend-detection scheme, not just wavelet-based detection—once we have computed  $\hat{\mu}_0$ ,  $\hat{\mu}_1$ , and  $\hat{\mu}_2$ , we can evaluate  $D_1^2$  and  $D_2^2$ , whichever we feel is appropriate. If the result is larger than the square of our presumed sensor resolution, we must detrend the  $\tilde{g}$  series before computing turbulence statistics to avoid a “significant” bias in the values.

## 6. Examples with Real Data

We have several data sets available with which to demonstrate wavelet-based trend detection. In August 1991, we collected over 60 hours of turbulence data with an ATI sonic anemometer/thermometer at the Sevilleta Long-Term Ecological Refuge, a semi-arid grassland near Socorro, New Mexico (Andreas 1995). The sonic was positioned 4 m above the surface and sampled at 10 Hz (i.e.,  $\Delta = 0.1$  s). We will use two short time series from this collection as our first examples. Both series come from the evening transition, when we would expect air temperature to begin falling. Hence, we look at sections of the sonic temperature record.

Figure 7 shows 300 s of sonic temperature data collected at about 8 p.m. local time. The measured temperature decreased during this interval but only by about 0.3°C. Our wavelet analysis, as described in Section 3, yields  $\hat{\mu}_0 = 21.27^\circ\text{C}$ ,  $\hat{\mu}_1 = -7.126 \times 10^{-6} ^\circ\text{C/s}$ , and  $\hat{\mu}_2 = -3.313 \times 10^{-6} ^\circ\text{C/s}^2$  for this series. From Eq. (5.10),  $D_2$  is then 0.09°C. If we take the resolution of the sonic thermometer to be 0.01°C, the reporting increment of the instrument, Eq. (5.10) requires that we remove the trend. Figure 7 also shows the quadratic trend that we removed and the detrended series. The original series has a standard deviation of 0.14°C; the detrended series, 0.12°C. Thus, we would have overestimated the standard deviation of this series by 17% if we had not detrended.

Figure 8 demonstrates why detrending is necessary. In it we plot the autocorrelation functions for the original and detrended series in Figure 7. Clearly, the autocorrelation function of the original series does not approach zero—even with a lag of 20 s—as we would expect of true surface-layer turbulence. As a result, the integral scale [see Eq. (3.22)]

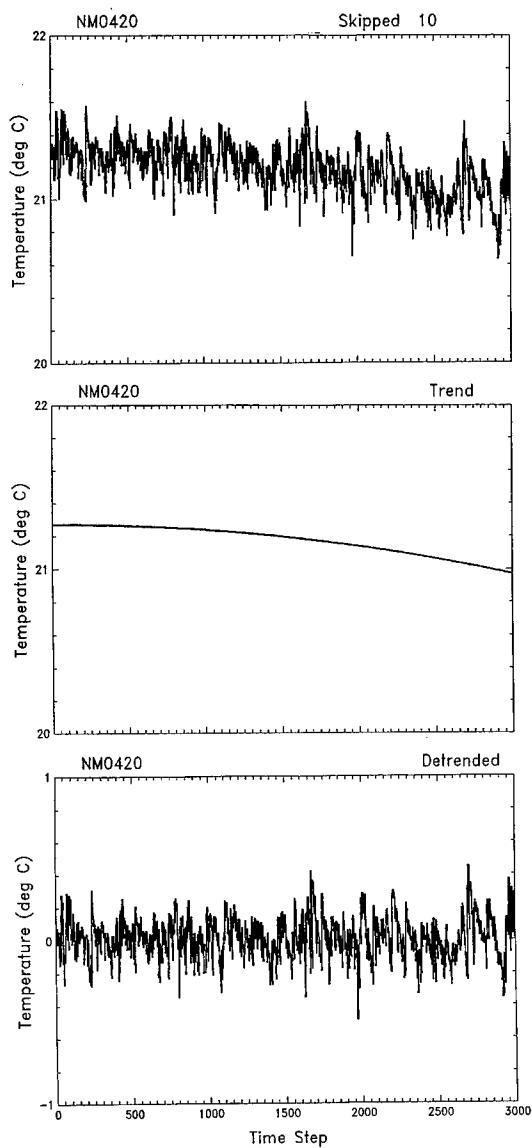


Fig. 7. Three hundred seconds of sonic thermometer data collected near 8 p.m. on 4 August 1991 at the Sevilleta Long-Term Ecological Refuge (top), the quadratic trend that we identified and removed (middle), and the detrended series (bottom). Each time step is 0.1 s.

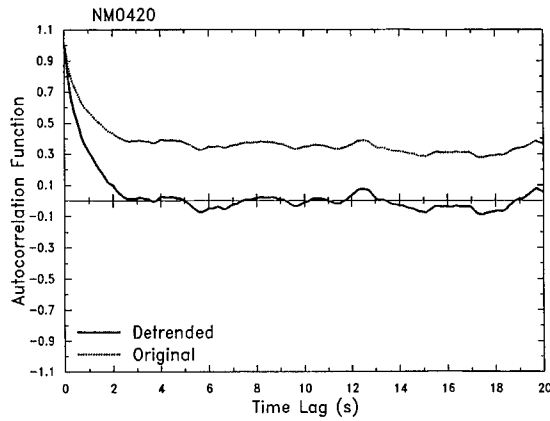


Fig. 8. Autocorrelation functions computed for the original and detrended series in Fig. 7. Integral scales of the original and detrended series are 7.5 and 0.79 s, respectively.

of the original time series is 7.5 s; while, for the detrended series, it is an order of magnitude shorter, 0.79 s.

Figure 9 shows a second Sevilleta sonic temperature series collected during another evening transition. Again, the original temperature series seems to be generally decreasing. Our wavelet analysis yields  $\hat{\mu}_0 = 25.25^\circ\text{C}$ ,  $\hat{\mu}_1 = 2.003 \times 10^{-3} \text{ }^\circ\text{C/s}$ , and  $\hat{\mu}_2 = -8.684 \times 10^{-6} \text{ }^\circ\text{C/s}^2$ . From Eq. (5.10),  $D_2$  is  $0.08^\circ\text{C}$ . Again, this is larger than the resolution in our temperature measurement,  $0.01^\circ\text{C}$ , so we must detrend. Figure 9 also shows that our detrending really levels the time series, which first increased in the mean and then decreased (Figure 9, middle panel).

The standard deviation of the original time series in Figure 9 is  $0.12^\circ\text{C}$ ; for the detrended series, it is  $0.09^\circ\text{C}$ . Therefore, using the original series to compute the standard deviation in temperature would have produced a value 33% higher than in the actual turbulence signal.

Lastly, the autocorrelation functions (Figure 10) of the original and detrended series shown in Figure 9 again emphasize the importance of detrending. The original series yields an autocorrelation function that goes to zero much more slowly than the one for the detrended series. From Figure 10 we compute the integral scale of the original series to be 8.1 s, while the integral scale of the detrended series is only 3.3 s.



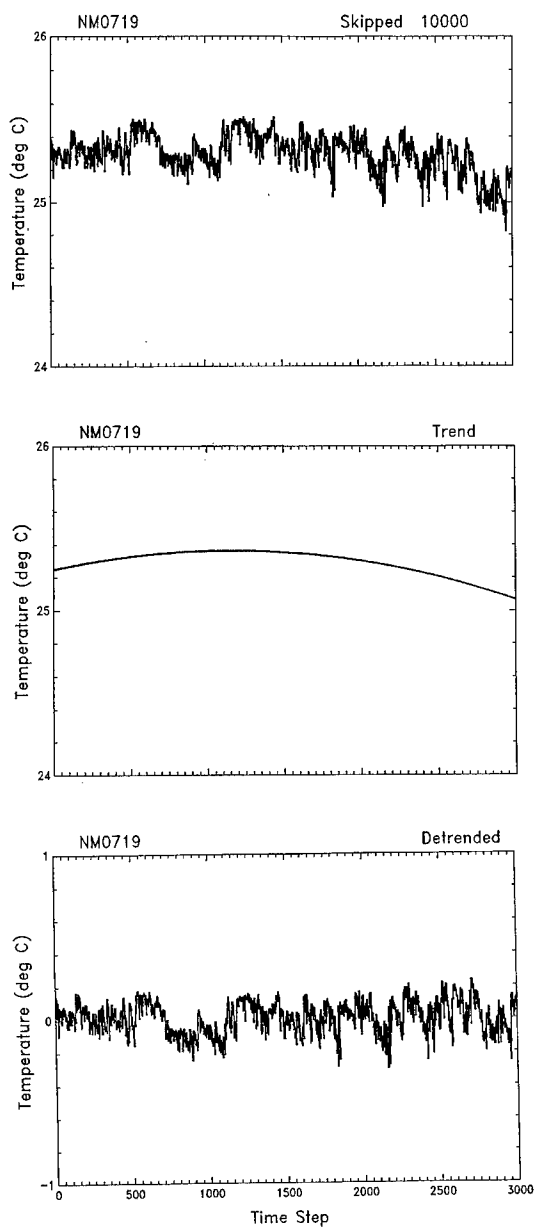


Fig. 9. As in Fig. 7, except these data were collected at about 7:15 p.m. local time on 7 August 1991.

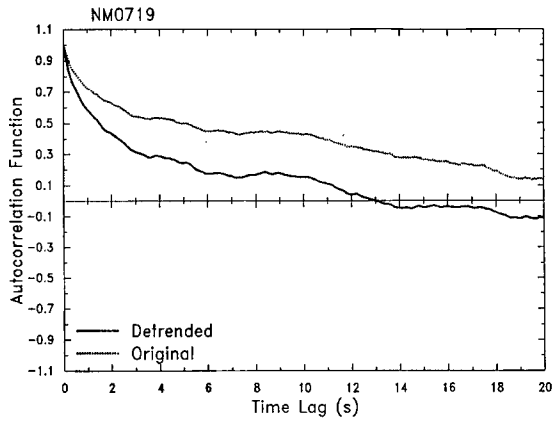


Fig. 10. Autocorrelation functions computed for the original and detrended series in Fig. 9. Integral scales for the original and detrended series are 8.1 and 3.3 s, respectively.

A second data set to which we apply our trend-detection scheme comes from Ice Station Weddell (ISW), a camp floating on the sea ice in the western Weddell Sea. On ISW we collected over 2000 hours of data with an R.M. Young (Traverse City, Michigan) propeller-vane anemometer mounted 5 m above the sea ice and sampled at 10 Hz (Andreas et al. 1992).

Figure 11 shows 5 minutes of wind speed data collected with the propeller-vane on 24 May 1992 at about 3 a.m. local time (Julian day 145, 0500 GMT) at an air temperature of  $-25^{\circ}\text{C}$ . Our wavelet analysis shows an almost linear trend in the series (Figure 11, middle panel);  $\hat{\mu}_0 = 7.93 \text{ m/s}$ ,  $\hat{\mu}_1 = -2.881 \times 10^{-3} \text{ m/s}^2$ , and  $\hat{\mu}_2 = 8.100 \times 10^{-7} \text{ m/s}^3$ . With a nominal resolution for the propeller-vane wind speed of 0.05 m/s (Andreas and Claffey 1995), Eq. (5.10) requires that we remove this trend since  $D_2 = 0.30 \text{ m/s}$ . The detrended series (Figure 11, bottom panel) has a standard deviation of 0.73 m/s, while the original series has a standard deviation of 0.78 m/s.

Figure 12 shows how detrending improves the behavior of the autocorrelation function for the series in Figure 11. The original series has an integral scale of 4.8 s; the detrended series, 2.9 s.

Finally, we close this section with an example of potential problems. Figure 13 shows another series of ISW propeller-vane wind speed data collected about 9 minutes after the series in Figure 11. Our wavelet calculations yield polynomial trend coefficients of  $\hat{\mu}_0 = 7.91 \text{ m/s}$ ,  $\hat{\mu}_1 = 1.411 \times 10^{-4} \text{ m/s}^2$ , and  $\hat{\mu}_2 = -7.068 \times 10^{-6} \text{ m/s}^3$ . These values yield  $D_2 = 0.18 \text{ m/s}$ ; Eq. (5.10) thus again requires detrending for a sensor resolution of 0.05 m/s.

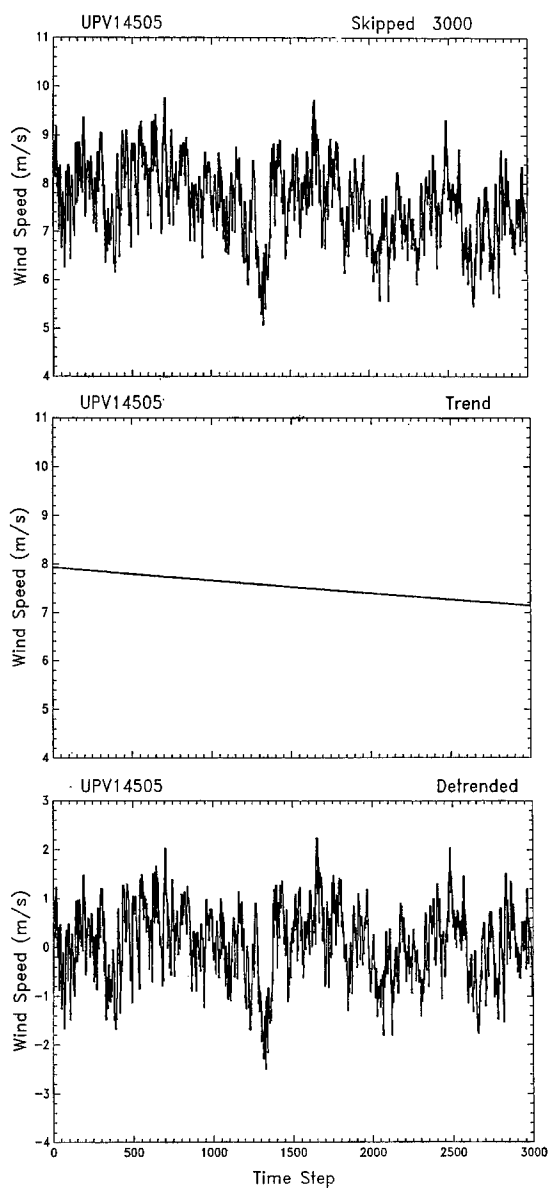


Fig. 11. Three hundred seconds of propeller-vane anemometer data collected at  $-25^{\circ}\text{C}$  on 24 May 1992 on Ice Station Weddell (top), the quadratic trend that we computed and removed (middle), and the detrended series (bottom).

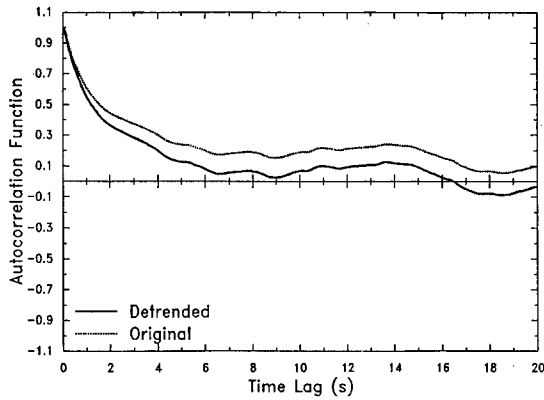


Fig. 12. Autocorrelation functions computed for the original and detrended series in Fig. 11. Integral scales for the original and detrended series are 4.8 and 2.9 s, respectively.

But the detrended series has a standard deviation of 0.81 m/s, while the original series has a standard deviation of only 0.80 m/s. Figure 14 shows that detrending also degrades the autocorrelation function. The original series has an integral scale of 4.7 s, while, for the detrended series, it is 5.1 s. From Eq. (5.8) it is possible to prove that the population variance of the turbulence alone ( $\sigma_g^2$ ) is never greater than the population variance of the turbulence signal riding on a trend ( $\sigma_g^2$ ). That is, the bracketed quantity on the right-hand side of Eq. (5.8) is never negative. Our computing sample variances that violate this constraint resulted because  $\hat{\mu}_0$ ,  $\hat{\mu}_1$ , and  $\hat{\mu}_2$  have random uncertainties. In the example in Figures 13 and 14, detrending introduced more variability into the resulting series through these uncertainties than the trend explained. Consequently, as a final criterion for trend removal, you must compare the variance of the original series with the variance of the detrended series. The latter must always be less, or the detrending is serving the opposite purpose for which it is intended.

## 7. Conclusions

We have described a new, wavelet-based method for detecting linear and quadratic trends in turbulence time series. Actually, though, there is nothing about our analysis that makes it specific to turbulence series. It would work just as well on a climatic time series or a series of Dow Jones averages.

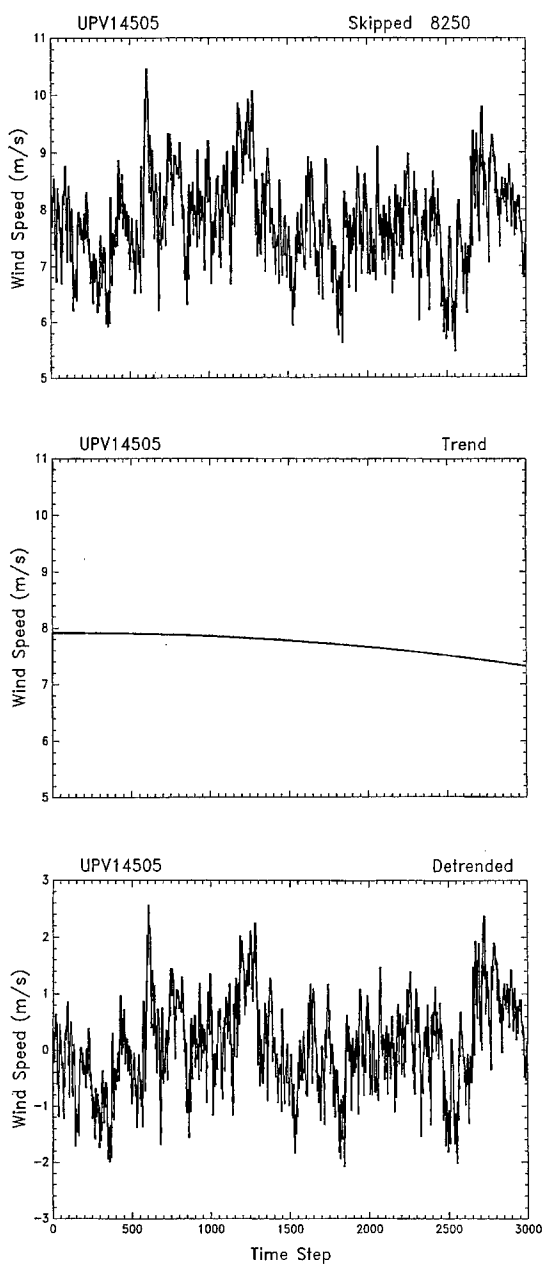


Fig. 13. As in Fig. 11, except about 9 minutes later.

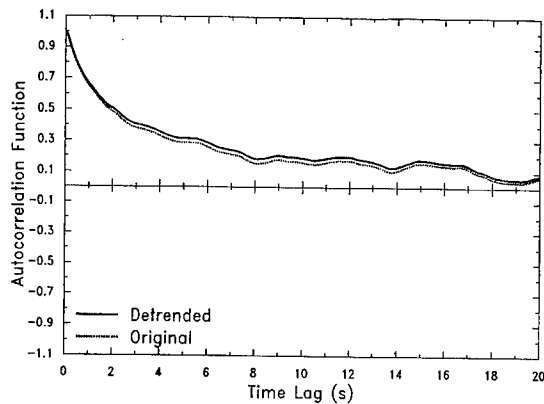


Fig. 14. Autocorrelation functions computed for the original and detrended series in Fig. 13. Integral scales for the original and detrended series are 4.7 and 5.1 s, respectively.

Fundamentally, wavelet trend detection works because the inverted Haar wavelet is a first-difference operator; it picks out the first-order coefficient of a quadratic trend. In turn, the elephant wavelet is a second-difference operator and, as such, picks out the second-order coefficient of the trend polynomial.

Although here we focused on polynomial terms only up to second order, it would be easy to extend our method to detect higher order trends by defining new (unnamed) wavelets. As we have shown, the wavelet for detecting a linear trend has lobes of  $-1$  and  $1$  with a multiplier of  $(2/L)^2$ . The wavelet for detecting a second-order trend has lobes of  $1$ ,  $-2$ , and  $1$  with a multiplier of  $(3/L)^3/2$ . By extension, the wavelet for detecting a third-order trend would have lobes of  $-1$ ,  $3$ ,  $-3$ , and  $1$  with a multiplier of  $(4/L)^4/6$ . And, in general, the wavelet to detect an  $n$ th-order trend would have lobes that are the coefficients of an  $n$ th-order, alternating binomial series with multiplier  $[(n+1)/L]^{n+1}/n!$ .

Wavelet-based trend detection is more efficient than its main competition—least-squares regression—in the sense that wavelet detection requires  $N$  fewer operations to quantify a linear trend and more than  $3N$  fewer operations to quantify a quadratic trend in a series of  $N$  samples (see Table 1). Least-squares regression, on the other hand, is more accurate than wavelet detection if the mean-square error of the estimators is the criterion. But the factor-of-two advantage in the accuracy of the  $\hat{\mu}_1$  and  $\hat{\mu}_2$  estimators that least-squares regression has over wavelet detection may not be worth the factor-of-two cost in computer operations.

A trend-detection scheme is not complete unless it provides guidance on when a trend is "significant" and therefore must be removed. We base our criterion for trend removal on the resolution,  $r$ , of the instrument used to collect the data. If the coefficients of the polynomial trend predict that the variance of the original series (with trend) and the variance of the actual turbulence series (no trend) differ by at least  $r^2$ , the trend will significantly bias the results and must be removed.

## 8. Acknowledgements

We thank M. G. Ferrick and D. G. Albert for valuable comments on the manuscript. The U.S. Department of the Army supported this work through Project 4A161102AT24; the U.S. National Science Foundation, with grants OPP-90-24544 and OPP-93-12642.

## 9. References

- Andreas, E.L., K.J. Claffey, A.P. Makshtas, and B.V. Ivanov, 1992: Atmospheric sciences on Ice Station Weddell. *Antarct. J. U. S.*, **27**(5), 115–117.
- Andreas, E.L., 1995: Statistics of surface-layer turbulence and evaluations of eddy-accumulation coefficients. Preprint volume, 11th Symposium on Boundary Layers and Turbulence, 27–31 March 1995, Charlotte, NC, American Meteorological Society, Boston, 106–109.
- Andreas, E.L., and K.J. Claffey, 1995: Air-ice drag coefficients in the western Weddell Sea: 1. Values deduced from profile measurements. *J. Geophys. Res.*, **100**, 4821–4831.
- Bendat, J.S., and A.G. Piersol, 1971: *Random Data: Analysis and Measurement Procedures*. Wiley-Interscience, 407 pp.
- Bevington, P.R., 1969: *Data Reduction and Error Analysis for the Physical Sciences*. McGraw-Hill, 336 pp.
- Collineau, S., and Y. Brunet, 1993: Detection of turbulent coherent motions in a forest canopy. Part I: Wavelet analysis. *Bound.-Layer Meteor.*, **65**, 357–379.
- Daubechies, I., 1988: Orthonormal bases of compactly supported wavelets. *Comm. Pure Appl. Math.*, **41**, 909–996.
- Farge, M., 1992: Wavelet transforms and their applications to turbulence. *Ann. Rev. Fluid Mech.*, **24**, 395–457.
- Gamage, N., and C. Hagelberg, 1993: Detection and analysis of microfronts and associated coherent events using localized transforms. *J. Atmos. Sci.*, **50**, 750–756.
- Gao, W., and B.L. Li, 1993: Wavelet analysis of coherent structures at the atmosphere-forest interface. *J. Appl. Meteor.*, **32**, 1717–1725.
- Haltiner, G.J., and R.T. Williams, 1980: *Numerical Prediction and Dynamic Meteorology*. John Wiley and Sons, 477 pp.
- Jenkins, G.M., and D.G. Watts, 1968: *Spectral Analysis and Its Applications*. Holden-Day, 525 pp.

- Katul, G.G., and M.B. Parlange, 1994: On the active role of temperature in surface-layer turbulence. *J. Atmos. Sci.*, **51**, 2181–2195.
- Lenschow, D.H., J. Mann, and L. Kristensen, 1994: How long is long enough when measuring fluxes and other turbulence statistics? *J. Atmos. Oceanic Technol.*, **11**, 661–673.
- Mahrt, L., and W. Gibson, 1992: Flux decomposition into coherent structures. *Bound.-Layer Meteor.*, **60**, 143–168.
- Mahrt, L., and J.F. Howell, 1994: The influence of coherent structures and microfronts on scaling laws using global and local transforms. *J. Fluid Mech.*, **260**, 247–270.
- Panofsky, H.A., and J.A. Dutton, 1984: *Atmospheric Turbulence: Models and Methods for Engineering Applications*. John Wiley and Sons, 397 pp.
- Papoulis, A., 1965: *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, 583 pp.
- Press, W.H., S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, 1994: *Numerical Recipes in FORTRAN: The Art of Scientific Computing*. Second Edition, Cambridge University Press, 963 pp.
- Priestley, M.B., 1981: *Spectral Analysis and Time Series*. Academic Press, 890 pp.
- Strang, G., 1989: Wavelets and dilation equations: A brief introduction. *SIAM Rev.*, **31**, 614–627.
- Treviño, G., 1982: A method for detrending correlated random data. *Applied Time Series Analysis*, O.D. Anderson and M.R. Perryman, Eds., North-Holland, 465–473.
- Turner, B.L., and M.Y. Leclerc, 1994: Conditional sampling of coherent structures in atmospheric turbulence using the wavelet transform. *J. Atmos. Oceanic Technol.*, **11**, 205–209.



### Appendix A: Mean-Square Error of Wavelet Trend Detection

Using Eq. (3.21) as the definition of the mean-square error, we want to evaluate here the accuracies of  $\hat{\mu}_0$ ,  $\hat{\mu}_1$ , and  $\hat{\mu}_2$  in the mse sense.

This analysis is easiest if we start with  $\hat{\mu}_2$ . From Eq. (3.21),

$$\text{MSE}[\hat{\mu}_2] = E[(\hat{\mu}_2 - \mu)^2]. \quad (\text{A1})$$

From Eq. (3.5), this becomes

$$\begin{aligned} \text{MSE}[\hat{\mu}_2] = & \\ B^2 E \left[ \iint_0^{L/3} ds dt \left[ g(s) - 2g\left(s + \frac{L}{3}\right) + g\left(s + \frac{2L}{3}\right) \right] \left[ g(t) - 2g\left(t + \frac{L}{3}\right) + g\left(t + \frac{2L}{3}\right) \right] \right]. & (\text{A2}) \end{aligned}$$

On multiplying the integrand out and defining the autocorrelation

$$C_g(s, t) \equiv E[g(s)g(t)], \quad (\text{A3})$$

we convert Eq. (A2) to

$$\begin{aligned} \text{MSE}[\hat{\mu}_2] = B^2 \iint_0^{L/3} ds dt & \left[ C_g(s, t) - 2C_g\left(s, t + \frac{L}{3}\right) + C_g\left(s, t + \frac{2L}{3}\right) \right. \\ & - 2C_g\left(s + \frac{L}{3}, t\right) + 4C_g\left(s + \frac{L}{3}, t + \frac{L}{3}\right) - 2C_g\left(s + \frac{L}{3}, t + \frac{2L}{3}\right) + C_g\left(s + \frac{2L}{3}, t\right) \\ & \left. - 2C_g\left(s + \frac{2L}{3}, t + \frac{L}{3}\right) + C_g\left(s + \frac{2L}{3}, t + \frac{2L}{3}\right) \right]. \end{aligned} \quad (\text{A4})$$

If the statistics of  $g$  are stationary,  $C_g(s, t)$  is an even function only of the difference  $\tau$  between  $s$  and  $t$ , where

$$\tau = t - s. \quad (\text{A5})$$

That is,

$$C_g(s, t) = C_g(s, s + \tau) = C_g(\tau) = C_g(-\tau). \quad (\text{A6})$$

Also, because of this symmetry, the double integration in Eq. (A4) reduces to a single integration over  $\tau$ ; for example, see Papoulis (1965, p. 325) or Panofsky and Dutton (1984, p. 62 ff.). That is,

$$\iint_0^L ds dt \Rightarrow 2L \int_0^L \left(1 - \frac{\tau}{L}\right) d\tau. \quad (\text{A7})$$

Therefore, Eq. (A4) becomes

$$\text{MSE}[\hat{\mu}_2] = \frac{2}{3}LB^2 \int_0^{L/3} d\tau \left(1 - \frac{3\tau}{L}\right) \times \left[6C_g(\tau) - 4C_g\left(\tau + \frac{L}{3}\right) - 4C_g\left(\tau - \frac{L}{3}\right) + C_g\left(\tau + \frac{2L}{3}\right) + C_g\left(\tau - \frac{2L}{3}\right)\right]. \quad (\text{A8})$$

Because  $C_g(\tau)$  approaches zero when  $\tau$  exceeds the integral scale  $\mathfrak{I}$  [see Eq. (3.22)] and because we assume that  $\mathfrak{I} \ll L$ , only the  $6C_g(\tau)$  term contributes to this integral. As a short demonstration of this, we evaluate two of the terms in Eq. (A8). The first term is

$$\int_0^{L/3} d\tau \left(1 - \frac{3\tau}{L}\right) [6C_g(\tau)] = 6\mathfrak{I}\sigma_g^2 \quad (\text{A9})$$

from Eq. (3.22). The  $3\tau/L$  term yields zero because where  $C_g$  is nonzero (for  $\tau \lesssim \mathfrak{I}$ ),  $3\tau/L \approx 0$ ; and where  $3\tau/L$  is large (for  $3\tau \approx L \gg \mathfrak{I}$ ),  $C_g$  is zero.

As another example, look at the second term in Eq. (A8);

$$\int_0^{L/3} d\tau \left(1 - \frac{3\tau}{L}\right) \left[4C_g\left(\tau + \frac{L}{3}\right)\right] = 4 \int_{L/3}^{2L/3} d\tau' \left(2 - \frac{3\tau'}{L}\right) C_g(\tau'), \quad (\text{A10})$$

where we have made the change of variables  $\tau' = \tau + \frac{L}{3}$ . But the right side of Eq. (A10) is zero because  $\mathfrak{I} \ll L$ .

By similar methods, we can show that the other three terms in Eq. (A8) also integrate to zero. Therefore,

$$\text{MSE}[\hat{\mu}_2] = \frac{2}{3}LB^2(6\mathfrak{I}\sigma_g^2). \quad (\text{A11})$$

Using Eq. (2.4) for B, we see

$$\text{MSE}[\hat{\mu}_2] = \frac{729}{L^4} \left(\frac{\mathfrak{I}}{L}\right) \sigma_g^2. \quad (\text{A12})$$

Next, we evaluate the mse of  $\hat{\mu}_1$ ;

$$\text{MSE}[\hat{\mu}_1] = E[(\hat{\mu}_1 - \mu)^2]. \quad (\text{A13})$$

From Eq. (3.12),

$$\begin{aligned} \text{MSE}[\hat{\mu}_1] &= A^2 E \left[ \int_0^{L/2} ds \int_0^{L/2} dt \left[ g\left(s + \frac{L}{2}\right) - g(s) \right] \left[ g\left(t + \frac{L}{2}\right) - g(t) \right] \right] \\ &\quad - 2ALE \left[ (\hat{\mu}_2 - \mu_2) \int_0^{L/2} dt \left[ g\left(t + \frac{L}{2}\right) - g(t) \right] \right] + L^2 E[(\hat{\mu}_2 - \mu_2)^2]. \end{aligned} \quad (\text{A14})$$

The third term is just

$$\text{3rd term} = L^2 \text{MSE}[\hat{\mu}_2] = \frac{729}{L^2} \left( \frac{\mathfrak{I}}{L} \right) \sigma_g^2. \quad (\text{A15})$$

Using Eq. (A3), (A6), and (A7), we can show that the first term in Eq. (A14) is

$$\text{1st term} = LA^2 \int_0^{L/2} \left( 1 - \frac{2\tau}{L} \right) \left[ 2C_g(\tau) - C_g\left(\tau - \frac{L}{2}\right) - C_g\left(\tau + \frac{L}{2}\right) \right] d\tau. \quad (\text{A16})$$

Again, because  $C_g(\tau)$  goes to zero rapidly when  $\tau$  exceeds  $\mathfrak{I}$ , and because  $\mathfrak{I} \ll L$ , only the first term in Eq. (A16) contributes to the integral. We find

$$\text{1st term} = LA^2 (2\mathfrak{I} \sigma_g^2) = \frac{32}{L^2} \left( \frac{\mathfrak{I}}{L} \right) \sigma_g^2. \quad (\text{A17})$$

From Eq. (3.5), the middle term in Eq. (A14) is

middle term =

$$-2ABLE \left[ \int_0^{L/3} ds \int_0^{L/2} dt \left[ g(s) - 2g\left(s + \frac{L}{3}\right) + g\left(s + \frac{2L}{3}\right) \right] \left[ g\left(t + \frac{L}{2}\right) - g(t) \right] \right]. \quad (\text{A18})$$

From Eq. (A3), this becomes

$$\begin{aligned} \text{middle term} = & -2ABL \int_0^{L/3} ds \int_0^{L/2} dt \left[ C_g\left(s, t + \frac{L}{2}\right) - C_g(s, t) - 2C_g\left(s + \frac{L}{3}, t + \frac{L}{2}\right) \right. \\ & \left. + 2C_g\left(s + \frac{L}{3}, t\right) + C_g\left(s + \frac{2L}{3}, t + \frac{L}{2}\right) - C_g\left(s + \frac{2L}{3}, t\right) \right]. \quad (\text{A19}) \end{aligned}$$

On using repeated variable transformations, we convert Eq. (A19) to

$$\begin{aligned} \text{middle term} = & -2ABL \left\{ \int_0^{L/3} ds \int_0^{L/2} dt \left[ C_g\left(s, t + \frac{L}{2}\right) - C_g(s, t) - 2C_g\left(s + \frac{L}{3}, t + \frac{L}{2}\right) \right. \right. \\ & \left. \left. + 2C_g\left(s + \frac{L}{3}, t\right) + C_g\left(s + \frac{2L}{3}, t + \frac{L}{2}\right) - C_g\left(s + \frac{2L}{3}, t\right) \right] \right\} \end{aligned}$$

$$\begin{aligned}
& + \iint_0^{L/6} ds dt \left[ C_g \left( s, t + \frac{5L}{6} \right) - C_g \left( s, t + \frac{L}{3} \right) - 2C_g \left( s + \frac{L}{3}, t + \frac{5L}{6} \right) \right. \\
& + 2C_g \left( s + \frac{L}{3}, t + \frac{L}{3} \right) + C_g \left( s + \frac{2L}{3}, t + \frac{5L}{6} \right) - C_g \left( s + \frac{2L}{3}, t + \frac{L}{3} \right) \\
& + C_g \left( s + \frac{L}{6}, t + \frac{5L}{6} \right) - C_g \left( s + \frac{L}{6}, t + \frac{L}{3} \right) - 2C_g \left( s + \frac{L}{2}, t + \frac{5L}{6} \right) \\
& \left. + 2C_g \left( s + \frac{L}{2}, t + \frac{L}{3} \right) + C_g \left( s + \frac{5L}{6}, t + \frac{5L}{6} \right) - C_g \left( s + \frac{5L}{6}, t + \frac{L}{3} \right) \right] \}. \quad (A20)
\end{aligned}$$

Invoking Eq. (A6) and (A7), we again convert the integrations in  $s$  and  $t$  to a single integration in  $\tau$ ;

$$\begin{aligned}
\text{middle term} = & -2ABL \left\{ \frac{2L}{3} \int_0^{L/3} d\tau \left( 1 - \frac{3\tau}{L} \right) \left[ C_g \left( \tau + \frac{L}{2} \right) - C_g(\tau) - 2C_g \left( \tau + \frac{L}{6} \right) \right. \right. \\
& + 2C_g \left( \tau - \frac{L}{3} \right) + C_g \left( \tau - \frac{L}{6} \right) - C_g \left( \tau - \frac{2L}{3} \right) \Big] \\
& + \frac{L}{3} \int_0^{L/6} d\tau \left( 1 - \frac{6\tau}{L} \right) \left[ C_g \left( \tau + \frac{5L}{6} \right) - 3C_g \left( \tau + \frac{L}{3} \right) - 2C_g \left( \tau + \frac{L}{2} \right) \right. \\
& \left. \left. + 3C_g(\tau) - C_g \left( \tau - \frac{L}{3} \right) + C_g \left( \tau + \frac{2L}{3} \right) + 2C_g \left( \tau - \frac{L}{6} \right) \right] \right\}. \quad (A21)
\end{aligned}$$

From Eq. (A9), we expect that the two  $C_g(\tau)$  terms will contribute to these integrals. We get

$$-\int_0^{L/3} d\tau \left( 1 - \frac{3\tau}{L} \right) C_g(\tau) = -3\sigma_g^2, \quad (A22)$$

and

$$3 \int_0^{L/6} d\tau \left( 1 - \frac{6\tau}{L} \right) C_g(\tau) = 33\sigma_g^2. \quad (A23)$$

Only the  $C_g\left(\tau - \frac{L}{6}\right)$  term in the first integral of Eq. (A21) makes an additional nonzero contribution to the middle term. That is,

$$\begin{aligned} \int_0^{L/3} d\tau \left(1 - \frac{3\tau}{L}\right) C_g\left(\tau - \frac{L}{6}\right) &= \int_{-L/6}^{L/6} d\tau' \left(\frac{1}{2} - \frac{3\tau'}{L}\right) C_g(\tau') \\ &= 2\left(\frac{1}{2}\right) \int_0^{L/6} d\tau' C_g(\tau') = 3\sigma_g^2. \end{aligned} \quad (A24)$$

Combining Eq. (A22–A24) in Eq. (A21) yields

$$\text{middle term} = -2ABL \left[ \frac{2L}{3}(-1+1) + \frac{L}{3}(3) \right] 3\sigma_g^2 = -\frac{108}{L^2} \left(\frac{\mathfrak{S}}{L}\right) \sigma_g^2. \quad (A25)$$

Finally, combining Eq. (A15), (A17), and (A25) in Eq. (A14) yields

$$\text{MSE}[\hat{\mu}_1] = \frac{653}{L^2} \left(\frac{\mathfrak{S}}{L}\right) \sigma_g^2. \quad (A26)$$

Lastly, we evaluate the mse of  $\hat{\mu}_0$ ;

$$\text{MSE}[\hat{\mu}_0] = E\left[(\hat{\mu}_0 - \mu_0)^2\right]. \quad (A27)$$

When we use Eq. (3.18),

$$\begin{aligned} \text{MSE}[\hat{\mu}_0] &= E\left[ \frac{1}{L^2} \int_0^L ds dt g(s)g(t) - (\hat{\mu}_1 - \mu_1) \int_0^L g(s)ds - \frac{2L}{3}(\hat{\mu}_2 - \mu_2) \int_0^L g(s)ds \right. \\ &\quad \left. + \frac{L^3}{3}(\hat{\mu}_1 - \mu_1)(\hat{\mu}_2 - \mu_2) + \frac{L^4}{4}(\hat{\mu}_1 - \mu_1)^2 + \frac{L^4}{9}(\hat{\mu}_2 - \mu_2)^2 \right]. \end{aligned} \quad (A28)$$

We see easily that the first term here is

$$\text{1st term} = \frac{2}{L} \int_0^L \left(1 - \frac{\tau}{L}\right) C_g(\tau) = 2\left(\frac{\mathfrak{S}}{L}\right) \sigma_g^2. \quad (A29)$$

The fifth term in Eq. (A28) is

$$\text{5th term} = \frac{L^4}{4} \text{MSE}[\hat{\mu}_1] = \frac{653}{4} \left( \frac{\mathfrak{I}}{L} \right) \sigma_g^2. \quad (\text{A30})$$

The sixth term is

$$\text{6th term} = \frac{L^4}{9} \text{MSE}[\hat{\mu}_2] = 81 \left( \frac{\mathfrak{I}}{L} \right) \sigma_g^2. \quad (\text{A31})$$

The second, third, and fourth terms in Eq. (A28) are more complex. We first evaluate the third term, which from Eq. (3.5) is

$$\text{3rd term} = -\frac{2}{3} \text{BLE} \left[ \int_0^L ds g(s) \int_0^{L/3} dt \left[ g(t) - 2g\left(t + \frac{L}{3}\right) + g\left(t + \frac{2L}{3}\right) \right] \right]. \quad (\text{A32})$$

On breaking the integral over  $s$  into three parts, transforming variables, and using Eq. (A3), we obtain

$$\begin{aligned} \text{3rd term} = & -\frac{2}{3} \text{BL} \int_0^{L/3} ds dt \left[ C_g(s, t) - 2C_g\left(s, t + \frac{L}{3}\right) + C_g\left(s, t + \frac{2L}{3}\right) \right. \\ & + C_g\left(s + \frac{L}{3}, t\right) - 2C_g\left(s + \frac{L}{3}, t + \frac{L}{3}\right) + C_g\left(s + \frac{L}{3}, t + \frac{2L}{3}\right) \\ & \left. + C_g\left(s + \frac{2L}{3}, t\right) - 2C_g\left(s + \frac{2L}{3}, t + \frac{L}{3}\right) + C_g\left(s + \frac{2L}{3}, t + \frac{2L}{3}\right) \right]. \quad (\text{A33}) \end{aligned}$$

As usual, invoking Eq. (A6) and (A7) converts this to

$$\begin{aligned} \text{3rd term} = & -\frac{2}{3} \text{BL} \int_0^{L/3} d\tau \left( 1 - \frac{3\tau}{L} \right) \left[ -C_g\left(\tau + \frac{L}{3}\right) + C_g\left(\tau + \frac{2L}{3}\right) \right. \\ & \left. - C_g\left(\tau - \frac{L}{3}\right) + C_g\left(\tau - \frac{2L}{3}\right) \right]. \quad (\text{A34}) \end{aligned}$$

Using the methods already employed in this appendix, we easily show that this term integrates to zero;

$$\text{3rd term} = 0. \quad (\text{A35})$$

Now look at the second term in Eq. (A28). From Eq. (3.12), we see that this is

$$\begin{aligned}
\text{2nd term} &= -AE \left[ \int_0^L ds g(s) \int_0^{L/2} dt \left[ g\left(t + \frac{L}{2}\right) - g(t) \right] \right] \\
&\quad + LE \left[ (\hat{\mu}_2 - \mu_2) \int_0^{L/2} ds g(s) \right]. \tag{A36}
\end{aligned}$$

The second term here is only a multiplication factor different from the third term in Eq. (A28), which we just showed was zero. Hence,

$$\text{2nd term} = -A \int_0^L ds \int_0^{L/2} dt \left[ C_g\left(s, t + \frac{L}{2}\right) - C_g(s, t) \right]. \tag{A37}$$

Again dividing the integration over  $s$  into two segments, transforming variables, and invoking Eq. (A6) and (A7), we convert Eq. (A37) to

$$\text{2nd term} = -AL \int_0^{L/2} d\tau \left( 1 - \frac{2\tau}{L} \right) \left[ C_g\left(\tau + \frac{L}{2}\right) - C_g(\tau) + C_g(\tau) - C_g\left(\tau - \frac{L}{2}\right) \right]. \tag{A38}$$

This also integrates to zero;

$$\text{2nd term} = 0. \tag{A39}$$

Finally, we tackle the fourth term in Eq. (A28). From Eq. (3.12), this is

$$\text{4th term} = \frac{1}{3} AL^3 E \left[ (\hat{\mu}_2 - \mu_2) \int_0^{L/2} dt \left[ g\left(t + \frac{L}{2}\right) - g(t) \right] \right] - \frac{1}{3} L^4 E \left[ (\hat{\mu}_2 - \mu_2)^2 \right]. \tag{A40}$$

We recognize  $\text{MSE}[\hat{\mu}_2]$  in the second term on the right-hand side of this. Thus, using Eq. (3.5),

$$\begin{aligned}
\text{4th term} &= \frac{1}{3} ABL^3 E \int_0^{L/3} ds \int_0^{L/2} dt \left[ g(s) - 2g\left(s + \frac{L}{3}\right) + g\left(s + \frac{2L}{3}\right) \right] \left[ g\left(t + \frac{L}{2}\right) - g(t) \right] \\
&\quad - \frac{1}{3} L^4 \text{MSE}[\hat{\mu}_2]. \tag{A41}
\end{aligned}$$

But we already evaluated the double integral in the first term of Eq. (A41) starting at Eq. (A18). Consequently,

$$\text{4th term} = \frac{1}{3}ABL^3(L\mathfrak{L}\sigma_g^2) - \frac{1}{3}L^4\left(\frac{729}{L^4}\frac{\mathfrak{L}}{L}\sigma_g^2\right) = -225\left(\frac{\mathfrak{L}}{L}\right)\sigma_g^2. \quad (\text{A42})$$

Finally, putting Eq. (A29–A31), (A35), (A39), and (A42) together in Eq. (A28) gives

$$\text{MSE}[\hat{\mu}_0] = \left[2 + 0 + 0 - 225 + \frac{653}{4} + 81\right]\frac{\mathfrak{L}}{L}\sigma_g^2 = 21.25\left(\frac{\mathfrak{L}}{L}\right)\sigma_g^2. \quad (\text{A43})$$



### Appendix B: Mean-Square Error of Least-Squares Trend Detection

The basis of least-squares trend detection is minimizing the total squared difference between the measured values and the fitted polynomial trend. In our application,  $\tilde{g}(t)$  is the measured series, and

$$\varepsilon^2 = \int_0^L dt [\tilde{g}(t) - \hat{\mu}_{0,LS} - \hat{\mu}_{1,LS}t - \hat{\mu}_{2,LS}t^2]^2 \quad (B1)$$

is the total squared difference between the measurements and the predicted trend.

Minimizing  $\varepsilon^2$  yields the least-squares estimators,  $\hat{\mu}_{0,LS}$ ,  $\hat{\mu}_{1,LS}$ , and  $\hat{\mu}_{2,LS}$ . We require

$$\frac{\partial \varepsilon^2}{\partial \hat{\mu}_{0,LS}} = 0 = -2 \int_0^L dt [\tilde{g}(t) - \hat{\mu}_{0,LS} - \hat{\mu}_{1,LS}t - \hat{\mu}_{2,LS}t^2], \quad (B2a)$$

$$\frac{\partial \varepsilon^2}{\partial \hat{\mu}_{1,LS}} = 0 = -2 \int_0^L dt t [\tilde{g}(t) - \hat{\mu}_{0,LS} - \hat{\mu}_{1,LS}t - \hat{\mu}_{2,LS}t^2], \quad (B2b)$$

$$\frac{\partial \varepsilon^2}{\partial \hat{\mu}_{2,LS}} = 0 = -2 \int_0^L dt t^2 [\tilde{g}(t) - \hat{\mu}_{0,LS} - \hat{\mu}_{1,LS}t - \hat{\mu}_{2,LS}t^2]. \quad (B2c)$$

Henceforth, we drop the subscript LS since this whole appendix treats only the least-squares estimators.

Equations (B2) integrate easily to yield

$$\hat{\mu}_0 + \frac{1}{2}\hat{\mu}_1L + \frac{1}{3}\hat{\mu}_2L^2 = \frac{1}{L} \int_0^L \tilde{g}(t) dt \quad (B3a)$$

$$\frac{1}{2}\hat{\mu}_0 + \frac{1}{3}\hat{\mu}_1L + \frac{1}{4}\hat{\mu}_2L^2 = \frac{1}{L^2} \int_0^L t \tilde{g}(t) dt, \quad (B3b)$$

$$\frac{1}{3}\hat{\mu}_0 + \frac{1}{4}\hat{\mu}_1L + \frac{1}{5}\hat{\mu}_2L^2 = \frac{1}{L^3} \int_0^L t^2 \tilde{g}(t) dt. \quad (B3c)$$

This set has an obvious matrix interpretation;

$$\begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{pmatrix} \begin{pmatrix} \hat{\mu}_0 \\ \hat{\mu}_1 L \\ \hat{\mu}_2 L^2 \end{pmatrix} = \begin{pmatrix} \frac{1}{L} \int_0^L \tilde{g}(t) dt \\ \frac{1}{L^2} \int_0^L t \tilde{g}(t) dt \\ \frac{1}{L^3} \int_0^L t^2 \tilde{g}(t) dt \end{pmatrix}. \quad (\text{B4})$$

We can invert the  $3 \times 3$  matrix readily to obtain solutions for  $\hat{\mu}_0$ ,  $\hat{\mu}_1$ , and  $\hat{\mu}_2$  (e.g., Bevington 1969, p. 134 ff.);

$$\hat{\mu}_0 = \frac{9}{L} \int_0^L \left( 1 - \frac{4t}{L} + \frac{10t^2}{3L^2} \right) \tilde{g}(t) dt, \quad (\text{B5})$$

$$\hat{\mu}_1 = -\frac{36}{L^2} \int_0^L \left( 1 - \frac{16t}{3L} + \frac{5t^2}{L^2} \right) \tilde{g}(t) dt \quad (\text{B6})$$

$$\hat{\mu}_2 = \frac{30}{L^3} \int_0^L \left( 1 - \frac{6t}{L} + \frac{6t^2}{L^2} \right) \tilde{g}(t) dt. \quad (\text{B7})$$

These then are the estimates for the polynomial trend coefficients that least-squares regression would yield. Let us investigate their statistical properties.

Look first at  $\hat{\mu}_0$ . In Eq. (B5), substitute Eq. (3.1) for  $\tilde{g}(t)$ ;

$$\hat{\mu}_0 = \frac{9}{L} \int_0^L dt \left( 1 - \frac{4t}{L} + \frac{10t^2}{3L^2} \right) [g(t) + \mu_0 + \mu_1 t + \mu_2 t^2]. \quad (\text{B8})$$

This readily integrates to

$$\hat{\mu}_0 = \frac{9}{L} \int_0^L dt \left( 1 - \frac{4t}{L} + \frac{10t^2}{3L^2} \right) g(t) + \mu_0. \quad (\text{B9})$$

Because  $E[g(t)] = 0$ , we see that

$$E[\hat{\mu}_0] = \mu_0. \quad (\text{B10})$$

That is,  $\hat{\mu}_0$  is an unbiased estimator of  $\mu_0$ .

Continue with  $\hat{\mu}_1$ . Again from Eq. (3.1),

$$\hat{\mu}_1 = -\frac{36}{L^2} \int_0^L dt \left( 1 - \frac{16t}{3L} + \frac{5t^2}{L^2} \right) [g(t) + \mu_0 + \mu_1 t + \mu_2 t^2]. \quad (\text{B11})$$

On integrating, we obtain

$$\hat{\mu}_1 = -\frac{36}{L^2} \int_0^L dt \left( 1 - \frac{16t}{3L} + \frac{5t^2}{L^2} \right) g(t) + \mu_1. \quad (\text{B12})$$

So, as before,

$$E[\hat{\mu}_1] = \mu_1; \quad (\text{B13})$$

$\hat{\mu}_1$  is an unbiased estimator of  $\mu_1$ .

Finally, for  $\hat{\mu}_2$ ,

$$\hat{\mu}_2 = \frac{30}{L^3} \int_0^L dt \left( 1 - \frac{6t}{L} + \frac{6t^2}{L^2} \right) g(t) + \mu_2, \quad (\text{B14})$$

and

$$E[\hat{\mu}_2] = \mu_2. \quad (\text{B15})$$

The estimator  $\hat{\mu}_2$  predicts  $\mu_2$  without bias.

Next, we look at the mean-square errors of the least-squares estimators. First, from Eq. (B9),

$$\text{MSE}[\hat{\mu}_0] = \frac{81}{L^2} E \left[ \iint_0^L ds dt \left( 1 - \frac{4s}{L} + \frac{10s^2}{3L^2} \right) \left( 1 - \frac{4t}{L} + \frac{10t^2}{3L^2} \right) g(s)g(t) \right]. \quad (\text{B16})$$

On multiplying this out, we obtain

$$\begin{aligned} \text{MSE}[\hat{\mu}_0] = \frac{81}{L^2} E \left[ \iint_0^L ds dt g(s)g(t) \left( 1 - \frac{4}{L}(s+t) \right. \right. \\ \left. \left. + \frac{10}{3L^2}(s^2+t^2) + \frac{16}{L^2}st - \frac{40}{3L^3}st(s+t) + \frac{100}{9L^4}s^2t^2 \right) \right]. \end{aligned} \quad (\text{B17})$$

Make the change of variables

$$\tau = t - s, \quad (\text{B18})$$

$$w = \frac{1}{2}(t + s). \quad (\text{B19})$$

This requires

$$\int_0^L ds dt \Rightarrow \int_{-L}^L d\tau \int_{\frac{|\tau|}{2}}^{\frac{L-|\tau|}{2}} dw. \quad (\text{B20})$$

Also, with Eq. (A3) and (A6), Eq. (B17) transforms to

$$\begin{aligned} \text{MSE}[\hat{\mu}_0] &= \frac{81}{L^2} \int_{-L}^L d\tau \int_{\frac{|\tau|}{2}}^{\frac{L-|\tau|}{2}} dw C_g(\tau) \\ &\times \left[ 1 - \frac{8w}{L} + \frac{68w^2}{3L^2} - \frac{7\tau^2}{3L^2} - \frac{80w^3}{3L^3} + \frac{20w\tau^2}{3L^3} + \frac{100}{9L^4} \left( w^4 - \frac{w^2\tau^2}{2} + \frac{\tau^4}{16} \right) \right]. \end{aligned} \quad (\text{B21})$$

On integrating over  $w$  in Eq. (B21), we obtain

$$\begin{aligned} \text{MSE}[\hat{\mu}_0] &= \frac{81}{L^2} \int_{-L}^L d\tau C_g(\tau) \\ &\times \left[ w - \frac{4w^2}{L} + \frac{68w^3}{9L^2} - \frac{7w\tau^2}{3L^2} - \frac{20w^4}{3L^3} + \frac{10w^2\tau^2}{3L^3} + \frac{100}{9L^4} \left( \frac{w^5}{5} - \frac{w^3\tau^2}{6} + \frac{w\tau^4}{16} \right) \right] \Bigg|_{\frac{|\tau|}{2}}^{\frac{L-|\tau|}{2}}. \end{aligned} \quad (\text{B22})$$

We realize now that when we integrate over  $\tau$ , any term in Eq. (B22) containing a  $\tau$  will integrate to zero because  $\Im \ll L$  and  $C_g(\tau)$  rapidly approaches zero when  $\tau$  exceeds  $\Im$ . Thus, in effect, we can take the integration limits for  $w$  to be 0 and  $L$ . With this simplification,

$$\text{MSE}[\hat{\mu}_0] = \frac{162}{L} \int_0^L d\tau C_g(\tau) \left[ 1 - 4 + \frac{68}{9} - \frac{20}{3} + \frac{20}{9} \right] = \frac{162}{9L} \int_0^L d\tau C_g(\tau). \quad (\text{B23})$$

Hence, from Eq. (3.22)

$$\text{MSE}[\hat{\mu}_0] = 18 \left( \frac{\Im}{L} \right) \sigma_g^2. \quad (\text{B24})$$

From Eq. (B12),

$$\text{MSE}[\hat{\mu}_1] = \left(\frac{36}{L^2}\right)^2 \mathbb{E} \left[ \iint_0^L ds dt \left(1 - \frac{16s}{3L} + \frac{5s^2}{L^2}\right) \left(1 - \frac{16t}{3L} + \frac{5t^2}{L^2}\right) g(s)g(t) \right]. \quad (\text{B25})$$

On multiplying this out, using the variable transformations given in Eq. (B18–B20), and invoking Eq. (A3), we obtain

$$\begin{aligned} \text{MSE}[\hat{\mu}_1] &= \frac{1296}{L^4} \int_{-L}^L d\tau \int_{\frac{|\tau|}{2}}^{\frac{L-|\tau|}{2}} dw C_g(\tau) \\ &\times \left[ 1 - \frac{32w}{3L} + \frac{346w^2}{9L^2} - \frac{83\tau^2}{18L^2} - \frac{160w^3}{3L^3} + \frac{40w\tau^2}{3L^3} + \frac{25}{L^4} \left( w^4 - \frac{w^2\tau^2}{2} + \frac{\tau^4}{16} \right) \right]. \quad (\text{B26}) \end{aligned}$$

Integrating over  $w$  yields

$$\begin{aligned} \text{MSE}[\hat{\mu}_1] &= \frac{1296}{L^4} \int_{-L}^L d\tau C_g(\tau) \\ &\times \left[ w - \frac{16w^2}{3L} + \frac{346w^3}{27L^2} - \frac{83w\tau^2}{18L^2} - \frac{40w^4}{3L^3} + \frac{20w^2\tau^2}{3L^3} + \frac{25}{L^4} \left( \frac{w^5}{5} - \frac{w^3\tau^2}{6} + \frac{w\tau^4}{16} \right) \right]_{\frac{|\tau|}{2}}^{\frac{L-|\tau|}{2}}. \quad (\text{B27}) \end{aligned}$$

But, as with Eq. (B22), we recognize that any term in the  $w$ - $\tau$  polynomial containing a  $\tau$  will integrate to zero when we do the next integration over  $\tau$ . Consequently, we can replace the limits

$$\frac{|\tau|}{2} \text{ and } L - \frac{|\tau|}{2}$$

with 0 and  $L$ . Equation (B27) then becomes

$$\text{MSE}[\hat{\mu}_1] = \frac{2592}{L^4} \int_0^L d\tau C_g(\tau) \left( 1 - \frac{16}{3} + \frac{346}{27} - \frac{40}{3} + 5 \right) = \frac{384}{L^2} \left( \frac{3}{L} \right) \sigma_g^2. \quad (\text{B28})$$

Finally, we evaluate the mse of  $\hat{\mu}_2$ . From Eq. (B14),

$$\text{MSE}[\hat{\mu}_2] = \frac{900}{L^6} \iint_0^L ds dt \left(1 - \frac{6s}{L} + \frac{6s^2}{L^2}\right) \left(1 - \frac{6t}{L} + \frac{6t^2}{L^2}\right) g(s)g(t). \quad (\text{B29})$$

Following exactly the same procedure as in the last two derivations, we get

$$\begin{aligned} \text{MSE}[\hat{\mu}_2] &= \frac{900}{L^6} \int_{-L}^L d\tau C_g(\tau) \\ &\times \left[ w - \frac{6w^2}{L} + \frac{16w^2}{L^2} - \frac{6w\tau^2}{L^2} - \frac{18w^4}{L^3} + \frac{9w^2\tau^2}{L^3} + \frac{36}{L^4} \left( \frac{w^5}{5} - \frac{w^3\tau^2}{6} + \frac{w\tau^4}{16} \right) \right] \frac{L - \frac{|\tau|}{2}}{\frac{|\tau|}{2}}. \quad (\text{B30}) \end{aligned}$$

Ignoring terms containing  $\tau$  here, as before, yields

$$\text{MSE}[\hat{\mu}_2] = \frac{1800}{L^5} \int_{-L}^L d\tau C_g(\tau) \left(1 - 6 + 16 - 18 + \frac{36}{5}\right) = \frac{360}{L^4} \left(\frac{3}{L}\right) \sigma_g^2. \quad (\text{B31})$$

## A TWO DIMENSIONAL POWER SPECTRAL STUDY FOR SOME NONSTATIONARY PROCESSES

GREGORY L. SMITH  
*Department of Mathematics*  
*Norfolk State University*  
*Norfolk, VA 23502*

A. G. MIAMEE  
*Department of Mathematics*  
*Hampton University*  
*Hampton, VA 23668*

Two dimensional spectral estimate of nonstationary processes is studied. Currently nonstationary phenomena are usually modeled and analysed as if they were stationary. The spectrum of a nonstationary process is two dimensional while that of a stationary process is one dimensional. The usual one dimensional spectrum and the more complete two dimensional spectrum of some nonstationary processes are compared. In addition, a random phase shift which as the theory shows should produce a stationary process, is introduced to our nonstationary process and the resulting stationary process is examined. These results are also compared with the method described above.

### 1 Introduction

In recent years interest has grown in nonstationary processes for modeling physical phenomena. This is due to the fact that most physical phenomena represent nonstationary behavior. Currently these phenomena are generally modeled as if they were stationary. This is because all aspects of the theory for representation, are quite well developed<sup>1</sup>. However, if these phenomena are to be better understood, they must be modeled as nonstationary. One aspect which distinguishes a nonstationary process from a stationary process is their power spectrum. Nonstationary processes have a two dimensional power spectrum while stationary processes have a one dimensional power spectrum. For nonstationary processes the theory is not complete, and some basic questions such as interpretation of their power spectrum still need to be investigated. In this paper, a two dimensional estimate of the power spectral density of a nonstationary processes will be discussed. The estimate then will be applied to some helicopter noise data which is clearly nonstationary. The results of this nonstationary analysis history will also be applied to our helicopter noise data.

The paper is organized as follows: Section 2 discusses the necessary background material, including the definition of three useful classes of nonstationary processes.

Section 3 discusses a method for determining the period of a periodically correlated process when the period is not known. The period of such a process must be known before one can attempt to produce a good estimate of the power spectrum. In Section 4, a method developed in<sup>9</sup> is used to obtain the power spectrum estimate. The bias and variability of the estimate are also discussed. Finally, in Section 5, the current method for analyzing nonstationary data is compared to our two dimensional spectral analysis developed in this paper. Conclusions are then made regarding the comparison of these methods.

## 2 Preliminaries and Background

A stochastic process or random process is a family of random variables  $\{X(\lambda) : \lambda \in \Lambda\}$ , where  $\Lambda$  is the index set of the parameter  $\lambda$ . Usually the index set is either the set  $Z$  of all integers in which case the process is called discrete or the set  $R$  of real numbers in which case the process is called continuous. Throughout this paper the procedures developed are for discrete processes. This is because the actual data used to produce the spectral estimate is discrete, a fact which is due to the necessity of sampling.

A stochastic process  $X(t)$  is called *stationary* if its covariance function

$$R(t, s) = E[X(t)X(s)], \quad t, s \in Z \quad (1)$$

satisfies the relation

$$R(t, s) = R(t + 1, s + 1),$$

for all  $t, s \in Z$ . A stochastic process  $X(t)$  is called *periodically correlated* (PC, in short) of period  $T$  if its covariance function satisfies

$$R(t, s) = R(t + T, s + T),$$

for all  $t, s \in Z$ , or if its covariance function written in the form

$$r_\tau(t) = R(t + \tau, t) = E[X(t + \tau)X(t)]$$

is periodic in  $t$  with period  $T$ . (A PC process with period  $T = 1$  is simply stationary). since  $r_\tau(t)$  is periodic in  $t$  with period  $T$ , one can write (for this and other results on PC processes see<sup>4</sup>):

$$r_\tau(t) = \sum_{k=0}^{T-1} R_k(\tau) \exp\left(\frac{2\pi i k t}{T}\right). \quad (2)$$

For convenience we usually extend the definition of functions  $R_k(\tau)$ ,  $k = 0, 1, \dots, T-1$ , to all integers by letting  $R_k(\tau) = R_{k+T}(\tau)$ . It is well-known that each  $R_k(\tau)$  has the representation

$$R_k(\tau) = \int_0^{2\pi} e^{-i\tau\lambda} dF_k(\lambda)$$



where each  $F_k(\cdot)$  is a measure on  $(0, 2\pi]$ . One can then write

$$r_\tau(t) = \int_0^{2\pi} \int_0^{2\pi} e^{-i(t+\tau)\lambda + i\theta} dF(\lambda, \theta)$$

with the two dimensional power spectrum  $F(\cdot, \cdot)$  of the process  $X(t)$  being

$$F(A, B) = \sum_{k=1-T}^{T-1} F_k \left( A \cap \left( B - \frac{2\pi k}{T} \right) \right),$$

(B-a is the set of all b-a's with b in B). In other words, spectrum  $F(\cdot, \cdot)$  is concentrated on  $2T-1$  straight line segments  $\lambda - \theta = 2\pi k/T, k = 1-T, \dots, T-1$ , contained inside the square  $T^2 = (0, 2\pi] \times (0, 2\pi]$ , with the measure  $F_k(\cdot)$  representing the mass of  $F(\cdot, \cdot)$  on the k-th segment. From the above discussion, taking  $T = 1$ , we see that the power spectrum of any stationary process is concentrated on the diagonal of the square  $T^2$ . Thus we recover the well-known one-dimensional spectrum of the stationary process, considered now as a two-dimensional spectrum. PC processes have been of interest because of their relatively simple structure which stems from their very close tie with stationary processes. In fact, it is well-known that: if  $X(t)$  is a PC process with period T. Then the T-variate process formed from consequent T blocks of  $X(t)$  is a multivariate stationary process. This has made it possible to employ properties of this associated stationary process and develop some structural properties of the PC process<sup>8</sup>. However, their close tie to stationary processes limits their modeling potential for nonstationary data. So, it is desirable to consider some richer classes of nonstationary processes. We now discuss one such class of nonstationary processes which was introduced by Hardin and Miamee<sup>9</sup>.

**Definition 1** A zero mean stochastic process  $X(t)$  is called covariance autocorrelated (CAR, in short) if there exists some finite set  $\{a_j : j = 1, 2, \dots, k\}$  of scalars such that

$$R(t, s) = \sum_{j=1}^K a_j R(t+j, s+j), \text{ for all } t, s \in Z. \quad (3)$$

Here are some examples of CAR processes

1. Recall that a process is stationary if its covariance function satisfies

$$R(t, s) = R(t+1, s+1)$$

and hence it is clearly CAR. In fact, in (1) one can take  $a_1$  to be 1 and the rest of  $a'_k s$  to be zero.

2. Similarly one can check that a PC process with period T is CAR, by taking  $k = T$ ,  $a_T = 1$  and the rest of  $a'_k s$  zero.

3. Let  $X(t)$  be a stationary process and let  $\alpha$  be a real number other than 1, then one can see that the following two processes

$$Y(t) = tX(t) \quad \text{and} \quad Z(t) = \alpha^t X(t)$$

are both CAR. For the first one, take  $k = 3$ ,  $a_1 = -3$ ,  $a_2 = -3$ , and  $a_3 = 1$ . Note that the CAR processes in example 3 are neither stationary nor PC. This is because, in each of these cases, the power of the process approaches infinity for large  $t$ 's, while for stationary and PC processes this quantity stays bounded.

4. Taking  $X(t)$  again to be a stationary process one can prove that the modulated processes

$$U(t) = (a + bt)tX(t) \quad \text{and} \quad V(t) = tatX(t)$$

are also CAR ( we omit the proof ).

Another important class of processes is that of harmonizable processes, which we briefly discuss below.

**Definition 2** A stochastic process  $X(t)$  is called harmonizable if its covariance function has the following harmonic (or Fourier ) representation

$$R(t, s) = \int_0^{2\pi} \int_0^{2\pi} e^{-i(t\theta - s\lambda)} dF(\theta, \lambda),$$

where the measure  $F$  defined on the square  $T^2$  is its spectral measure.

In other words, we say a random process  $X(t)$  is harmonizable if the double Fourier transform of its covariance function

$$f(\lambda, \theta) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} ds R(t, s) e^{-i(\lambda t - \theta s)}, \quad (4)$$

exist. If this is the case, then  $f(., .)$  is called the power spectral density of the process  $X(t)$ . For a stationary process, replacing  $R(t, s)$  by  $r(t-s)$  and substituting  $\tau$  for  $t-s$ , (4) simplifies to

$$f(\lambda, \theta) = f\left(\frac{\lambda + \theta}{2}\right) \delta(\lambda - \theta),$$

where  $\delta(.)$  is the Dirac delta function, and

$$f(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} R(\tau) e^{-i\lambda\tau} d\tau.$$

This results in a one dimensional spectral density.

We showed that any stationary process is harmonizable and its (two dimensional) spectral measure is concentrated on the diagonal  $D = \{(\lambda, \theta) : \lambda = \theta\}$  of the square  $T^2$ . One can similarly show that any PC process with period  $T$  is harmonizable with

its spectral measure being concentrated on  $2T-1$  equidistant line segments parallel to the diagonal of that square, i.e. on the line segments

$$D_k = \{(\theta, \lambda) : \theta = \lambda + 2\pi/k, k = -T+1, \dots, T-1\}$$

Now we present the following theorem which is essential for any application of CAR processes ( for its proof and other properties of CAR processes the reader is referred to <sup>6</sup>).

**Theorem 1** *The power spectrum of any harmonizable CAR process is concentrated on a finite number of straight line segments parallel (but not necessarily equidistant from) the main diagonal of the square  $T^2$ .*

### 3 Estimating period of a PC process

In this section, a method will be examined for determining the period of a PC process from its given data. The method entails finding the lines of spectral support for the PC process and then to use the distance between these lines to determine the period.

The discrete data we analyze here was taken from an acoustic time history produced by a helicopter that was fixed with respect to the observer. The passing blades from an isolated main rotor or isolated tail rotor will produce a periodic sound pressure time history. This time history will have the same period as that of the passing blades which produced them. The history produced by the main or tail rotor rotating alone represents a PC process which covariance function has the same period as the time history. Recall that covariance function of a PC process is periodic with a period  $T$  and has an associated frequency  $f$ . It is this period that we want to find.

Hurd<sup>6</sup> developed a useful technique for determining the spectral support of a nonstationary process. The technique consists of first producing the discrete Fourier transform  $\bar{X}_k$  from the sampled data. Then products  $\bar{X}_p \bar{X}_q$  of these transforms are obtained and plotted in the  $(p, q)$  plane. Subsets of these products are then summed along the diagonal and normalized with respect to the main diagonal. That is, a spectral coherence is produced at coordinates  $(p, q)$  where  $p$  and  $q$  correspond to the  $2p/N\Delta t$  and  $2q/N\Delta t$  frequencies respectively ( the time between subsequent samples). The coherence is produced by

$$|r(p, q, M)|^2 = \frac{|\sum_{m=0}^{M-1} \bar{X}_p \bar{X}_q|^2}{\sum_{m=0}^{M-1} |\bar{X}_{p+m}|^2 \sum_{m=0}^{M-1} |\bar{X}_{q+m}|^2}.$$

This coherence is used to determine which points over the array being considered have significant values. This is determined by choosing a threshold value and plotting points  $(p, q)$  for which the coherence exceeds the threshold. If the process is PC, the plotted points  $(p, q)$  should produce a graph ( see for example figure 3 ) of dots along lines parallel to the diagonal. The separation between these parallel lines can be used to see whether the process is PC or not, and to find the period of our process.

This technique is based on the theoretical result discussed in Section 1 which says that the spectral support for a PC process with period  $T$  is on  $2T-1$  equidistant line segments parallel to the main diagonal of the square  $T^2$ . That is on the line segments

$$D_k = \{(\lambda, \theta) : \lambda = \theta + 2\pi k/T\}, k = 0, 1, \dots, (T-1).$$

Taking the case with  $k=1$ , the spacing of these lines is given by

$$p - q = \frac{2\pi}{T}$$

letting  $T = n\Delta t$ , produces

$$p - q = \frac{2\pi}{n\Delta t}$$

The difference  $p-q$  is found from the graph. Thus the number  $n$  of time intervals  $\delta t$  needed to produce the period of the process can be calculated directly. After the period has been found, our next task is to estimate the power spectral density.

#### 4 Power Spectral Estimation

A random variable  $\bar{\lambda}$  which is an estimate for an unknown parameter  $\lambda$  is called unbiased if

$$E(\bar{\lambda}) = \lambda,$$

In addition, an estimate is of interest if its uncertainty is as small as possible. The uncertainty of an estimate is measured by its standard deviation. That is,

$$\sigma = (E[\bar{\lambda} - \lambda]^2)^{\frac{1}{2}}$$

Therefore, the usual requirements are that the estimate be unbiased and have the smallest possible standard deviation  $\sigma$ .

For stationary processes there are two well known techniques for spectral estimation which are the Blackman-Tukey and the finite Fourier transform techniques. Here, to produce an estimate for the power spectral density, we have chosen the finite Fourier transform technique. This method consists of taking the discrete Fourier transform of sampled data and using the transformed data to produce a spectral estimate. Consider a stationary process  $X(t)$  of which the (generalized) Fourier transform is given by

$$\bar{X}(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(t) e^{-i\omega t} dt.$$

So, its covariance function is given as

$$E[\bar{X}(\lambda)\bar{X}(\theta)] = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} ds R(t-s) e^{-i(\lambda t - \theta s)}.$$

Changing variables  $(t + s)/2$  and  $s - t$  to  $t$  and  $\tau$ , respectively, produces

$$E[\bar{X}(\lambda)\bar{X}(\theta)] = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} d\tau R(\tau) e^{\frac{i(\lambda+\theta)\tau}{2}} \int_{-\infty}^{\infty} dt e^{-i(\lambda-\theta)t} = f\left(\frac{\omega + \theta}{2}\right) \delta(\omega - \theta).$$

Therefore,

$$E[\bar{X}(\lambda)]^2 = f(\lambda)$$

In practical situations only a single sample function of finite length of a random process  $X(t)$  is available. And based on the relationship just obtained, a class of power spectral estimates

$$\hat{f}(\lambda) = W_s |\bar{X}_f(\lambda)|^2$$

is introduced where

$$\bar{X}(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} n(t) X(t) e^{i\lambda t} dt \quad (5)$$

is the Fourier transform of the data as seen through a data window  $n(t)$ . This data window is a real valued function that is zero for  $t < 0$  and  $t > T$ , so that unavailable data are not required. And  $W_s$  is a correction factor, due to the presence of the window. The estimate  $\hat{f}(\lambda)$  for a fixed  $\lambda$  is a random variable with mean

$$E[\hat{f}(\lambda)] = \frac{w_s}{4\pi^2} \int_{-\infty}^{\infty} dt n(s) n(t) R(t-s) e^{-i\lambda(s-t)}$$

Furthermore,

$$E[\hat{f}(\lambda)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} dt \frac{W_s}{2\pi} \int_{-\infty}^{\infty} d\tau n(t + \tau/2) n(t - \tau/2) R(\tau) e^{-i\lambda\tau},$$

with  $t - \tau/2$  and  $t + \tau/2$  being substituted for  $s$  and  $t$ , respectively. Here

$$u(\tau) = \frac{W_s}{2\pi} \int_{-\infty}^{\infty} n(t + \tau/2) n(t - \tau/2) dt$$

is a lag window satisfying the following conditions:

1.  $u(0) = 1$ , for preserving power
2.  $u(\tau) = u(-\tau)$  which makes  $\hat{f}(\lambda)$  real
3.  $u(\tau) = 0$  for  $|\tau| > T$ .

For the first condition to hold we arrive at the requirement,

$$W_s = \frac{2\pi}{\int_{-\infty}^{\infty} n^2(t) dt}.$$

Therefore, the estimate becomes

$$\hat{f}(\omega) = \frac{2\pi |\bar{X}(\omega)|^2}{\int_{-\infty}^{\infty} n^2(t) dt}.$$

The second condition, namely  $u(\tau)$  being even, is obviously satisfied. The third condition is also satisfied since  $u(\tau)$  is the convolution of two data windows that are only nonzero in the interval  $(0, T)$ .

The estimate developed above is equivalent in expectation to a class of estimates developed by Blackman and Tukey, in which

$$\hat{f}(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} u(\tau) \hat{R}(\tau) e^{-i\lambda\tau} d\tau.$$

Here  $\hat{R}(\tau)$  is an estimate of the covariance function of the process and  $u(\tau)$  is a lag window as described above. The expectation of the estimate is

$$E[\hat{f}(\lambda)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} u(\tau) R(\tau) e^{-i\lambda\tau} d\tau.$$

This mean spectral estimate can be shown to equal the convolution of the actual spectral density with a "spectral window" which is merely the Fourier transform of the lag window. Since the covariance is an even function, one can see that

$$E[\hat{f}(\lambda)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} u(\tau) R(\tau) \cos \lambda\tau d\tau = f(\lambda) - \frac{1}{2\pi} \int_T^{\infty} \cos \lambda\tau d\tau$$

for the case of a "boxcar" lag window,  $u(\tau) = 1$ . Therefore, the estimate for the power spectral density is biased, but it becomes unbiased as  $T \rightarrow \infty$ .

In a similar fashion to the first technique, we suggest the following procedure for estimating the power spectral density of nonstationary processes. The power spectral density for a nonstationary process can be written as

$$\hat{f}(\lambda, \theta) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} ds \int_{-\infty}^{\infty} dt R(t, s) e^{-i(\lambda s - \theta t)} = E[\bar{X}(\lambda) \bar{X}(\theta)],$$

where

$$\bar{X}(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(t) e^{-i\lambda t} dt.$$

Here  $f(\lambda, \theta)$  can not be simplified to  $f(\lambda)$  since the covariance function  $R(t, s)$  depends on both variables  $t$  and  $s$ . Therefore, from a sample function of length  $T$  of a nonstationary process  $X(t)$ , a class of power spectral estimates  $\hat{f}(\lambda, \theta) = W_s \bar{X}(\lambda) \bar{X}(\theta)$  can be introduced when  $\bar{X}(\lambda)$  is exactly as introduced for the stationary case in (5). The mean of the estimate is

$$E[\hat{f}(\lambda, \theta)] = \frac{W_s}{4\pi^2} \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} ds n(t) n(s) R(t, s) e^{-i(\lambda t - \theta s)}.$$

The above analysis for the estimate being bias when  $\lambda = \theta$  in the stationary case can be seen to hold when  $\lambda$  is not equal to  $\theta$ . It can be shown that the spectral estimate of the finite Fourier transform, in the stationary case, is essentially a chi-square random variable with two degrees of freedom. Therefore, variability of the estimate can be reduced by breaking the estimate into  $N_B$  blocks of length  $T_B$  such that  $N_B T_B = T$ .

A spectral estimate  $f_j(\lambda)$ , for  $j = 1, 2, \dots, N_B$  will then be taken over each block. If the blocks are assumed to be independent, the average of the block estimates,

$$\bar{f}(\lambda) = \frac{1}{N_B} \sum_{j=0}^{N_B} \hat{f}_j(\lambda)$$

is essentially a chi-square random variable with  $K = 2N_B$  degrees of freedom. The variability of such estimate is intimately linked to its resolution. Full resolution refers to two sinusoids of the same amplitude being completely distinguishable when viewed through the spectral window function in the frequency domain. For the finite Fourier transform technique, full resolution requires that frequencies be roughly separated by

$$\Delta\lambda = \frac{2\pi}{T} \quad \text{or} \quad \Delta f = \frac{1}{T}.$$

Now, by breaking the data into  $T_B$  blocks the effective length has changed from  $T$  to  $T_B$ . Thus, the bandwidth of the estimate decreases to

$$\Delta f = \frac{1}{T_B}$$

Since,  $K = 2N_B = \frac{2T}{T_B}$  we have  $K = 2T\Delta f$  which shows the tradeoff between variability and frequency resolution. The estimate

$$\hat{f}(\lambda) = W_s \bar{X}(\lambda) \bar{X}(\lambda)$$

of the stationary case was a chi-square random variable but the estimate

$$\hat{f}(\lambda, \theta) = W_s \bar{X}(\lambda) \bar{X}(\theta)$$

in the case of a PC process was not so. However, the latter estimate similarly reduces in variability, when the process is blocked in integer multiples of the period. i.e.  $T_B = nT$ ,  $T$  being the period.

## 5 Spectral Analysis of Helicopter Noise

We will first examine the correlation of the acoustic pressure time history  $X(t)$  from an isolated helicopter rotor.  $X(t)$  is a discrete process representing the time history sampled at discrete intervals. It is this sample data which is used to produce our

spectral estimate. Due to the rotating nature of the helicopter blades we can assume that the time history is periodically correlated. One might further assume that this time history is actually doubly periodically correlated (DPC). If this was the case, then its covariance function  $R(s,t)$  will be periodic in both  $t$  and  $s$ . The correlation function being periodic in  $t$  can then be written as

$$R(s,t) = \sum_{k=0}^{T-1} R_k(t) e^{\frac{2\pi i s k}{T}} \quad (6)$$

for all  $s \in Z$ . Now since  $R(s,t)$  is also periodic in  $t$ , i.e.  $R(s,t+T) = R(s,t)$ , we can write

$$\sum_{k=0}^{T-1} R_k(t) e^{\frac{2\pi i s k}{T}} = \sum_{k=0}^{T-1} R_k(t+T) e^{\frac{2\pi i s k}{T}}, \text{ for all } s, t \in Z.$$

For each fixed  $t$  we have

$$0 = \sum_{k=0}^{T-1} [R_k(t+T) - R_k(t)] e^{\frac{2\pi i s k}{T}}, \text{ for all } s \in Z.$$

which implies that

$$R_k(t+T) = R_k(t), \text{ for all } k \in Z.$$

Now, since  $t$  was arbitrary,  $R_k(t)$  is periodic for each  $k$ , and we can write

$$R_k(t) = \sum_{j=0}^{T-1} a_{kj} e^{\frac{2\pi i t j}{T}}$$

Substituting this into (6) and simplifying the result we get

$$R(s,t) = \sum_{k=0}^{T-1} \sum_{j=0}^{T-1} a_{kj} e^{\frac{2\pi i (s k + t j)}{T}} \quad (7)$$

This means that the spectrum of a DPC processes is supported on  $T^2$  points (see figure 1)

$$(2\pi k/T, 2\pi j/T) : j, k = 0, 1, \dots, T-1 \quad (8)$$



However as we will show later in figure 3 the spectral support of our data is not on lattice points and hence our process is not DPC.

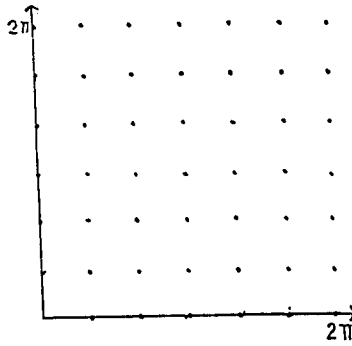


Figure 1. Spectral Support for DPC Processes

Now we can try to get a representation for the process  $X(t)$  itself. To do this one should first verify that the matrix  $A = [a_{kj}]$  is positive definite (we omit the proof). This done, it is well-known that  $A = [a_{kj}]$  is covariance matrix of some zero-mean Gaussian random vector  $(Y_0, Y_1, \dots, Y_{T-1})$ . That is to say

$$a_{kj} = E[Y_k Y_j],$$

for all  $k, j = 0, 1, \dots, T-1$ . Therefore, we can write (7) as

$$R(s, t) = \sum_{k=0}^{T-1} \sum_{j=0}^{T-1} E(Y_k Y_j) e^{\frac{2\pi i(k s + j t)}{T}}$$

or

$$R(s, t) = E\left[\left(\sum_{k=0}^{T-1} e^{\frac{2\pi i s k}{T}} Y_k\right) \left(\sum_{j=0}^{T-1} e^{\frac{2\pi i t j}{T}} Y_j\right)\right]$$

This means that  $X(t)$  can be taken to be

$$X(t) = \sum_{k=0}^{T-1} e^{\frac{2\pi i t k}{T}} Y_k, \text{ for all } t \in \mathbb{Z}.$$

This in particular shows  $X(t) = X(t+T)$ , for all  $t$ , which means the process  $X(t)$  is periodic. Now, since  $X(t)$  is periodic, its prediction is easy once its period is known.

### 5.1 Two-dimensional analysis

Our aim in the rest of this section is to use the results discussed above to analyze our helicopter noise. The standard approach of studying the data is to treat it as if it were stationary and hence produce and study its one dimensional power spectrum. However, in actuality the data produced from the isolated tail rotor or isolated main rotor is not stationary but PC. So it takes a two-dimensional power spectrum to adequately study such data. Furthermore, the combined main and tail rotor data is neither PC nor stationary and it is likely to be CAR. Hence this data, again, requires a two dimensional spectral analysis.

In Section 3, an estimate was developed for a nonstationary process. We will use this estimate to produce two-dimensional spectral estimates for the data. The technique consist of estimating the spectrum over a region for values of  $\lambda$  and  $\theta$  and case of the isolated main and tail rotors blocking was done as a multiply of the period of process. This was done because of the characteristics of PC processes. To insure that these points are actually viewed, a length of data which is a multiple of the period of the process must be used when the data is Fourier transformed.

The data used to produce the spectral estimates in this section Sikorsky Aircraft's Basic Model Test Rig was used, and the details of which can be found in<sup>9</sup>. The data was taken from several locations around the helicopter model (figure 2). By analyzing data from several different positions, the noise pattern produced could be used to determine how the noise is radiated in different directions and to identify common characters.

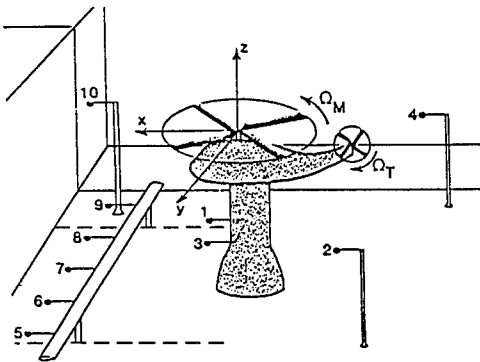


Figure 2. Diagram of Microphone Locations

In producing the two dimensional spectrum for the isolated tail rotor, a Fortran program was written to produce values of the 450 Hz which was determined by graphed data ( see figure 3, where lines for the shaft frequency are also present) produced by another computer program based on our discussion in Section 2. The test conditions. The passing blade produces a corresponding periodic acoustic pressure time history with a period of 450 Hz. After the data was inputted into our program for producing a two-dimensional spectrum, a set of values were given in the output for the spectral estimate at frequencies  $\lambda$  and  $\theta$  going from zero frequency to a chosen upper frequency (for us 4800 Hz). We viewed the data and decided that due to the background noise at low frequencies the scaling of the graph did not allow enough of the power at the fundamental frequency and at harmonics of the tail rotor period to be shown. Therefore, we decided to filter the data using a Chebyshev digital filter. After we removed all frequencies below 250 Hz, the graph of the output data revealed more detail of the power at the fundamental frequencies and its harmonics (see Figures 4 and 5)

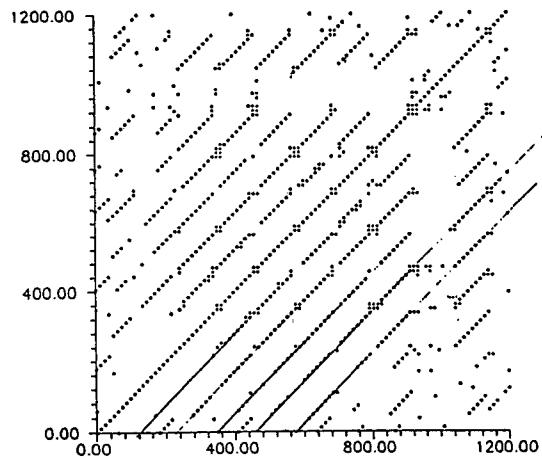


Figure 3. Spectral Coherence for Isolated Tail Rotor Data

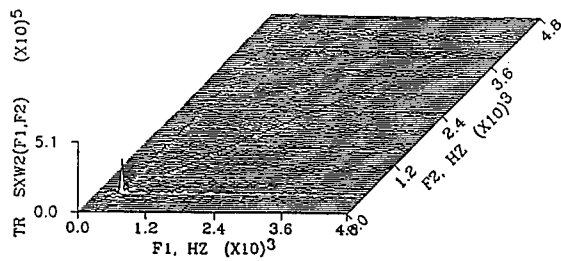


Figure 4. Tail Rotor Spectrum Microphone 2

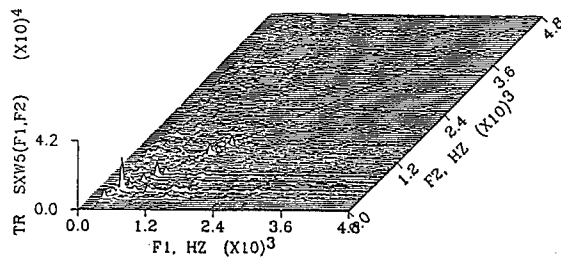


Figure 5. Tail Rotor Spectrum Microphone 5

The graphs show most of the power for the tail rotor at the fundamental frequency (450 Hz) and the first and second harmonics (900 and 1350 Hz respectively). There are components off the main sinusoids at frequencies  $\lambda$  and  $\theta$  where  $\lambda \neq \theta$ . As for the data sampled from the isolated main rotor time history, this data is periodic with a frequency of 95 Hz. The first graphs we produced from this data showed high power levels at around 95 Hz. to filter frequencies below 250 Hz out of this data also. This resulted in the second harmonic of the main rotor noise (285 Hz) being dominant. At microphone 5 the background tonal noise between 800 and 1300 Hz as reported in<sup>11</sup> is apparent ( see figures 6 and 7 for the spectral produced from data taken by microphones 2 and 5).

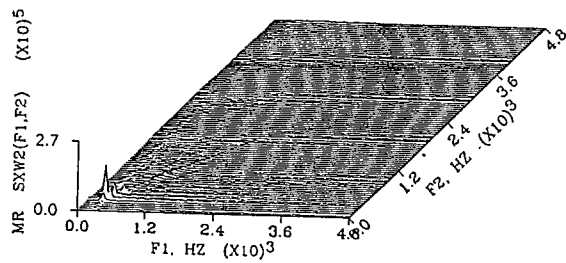


Figure 6. Main Rotor Spectrum Microphone 2

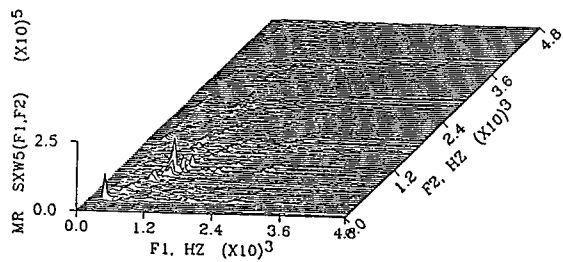


Figure 7. Main Rotor Spectrum Microphone 5

The spectrum of the combined noise was of interest due to the fact that if noises for the isolated main and tail rotors were independent, the theory suggest that the spectrum for the combined noise is just the sum of the individual spectra. i.e. *spectrum of combined noise must be the same as combined spectrum of noises*. The spectrum of the combined noise which is the noise produced when both the main and tail rotors are working was of interest due to the fact that if these noises are independent, the theory shows that the spectrum for the combined noise is just the sum of the spectra of the isolated main and tail rotors. We decided to determine the dependence of these noises by comparing the spectrum of the combined noises (see figures 8 and 9) with the added spectrum of the noises of the isolated tail rotor and main rotor.

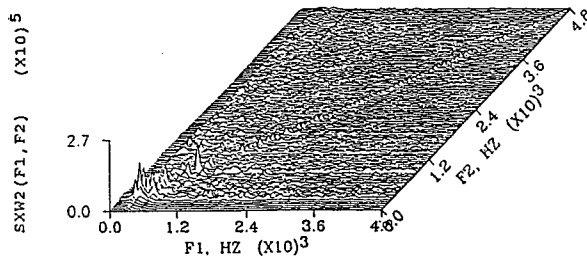


Figure 8. Combined Spectrum Microphone 2

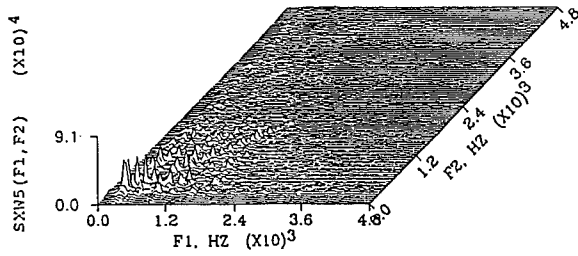


Figure 9. Combined Spectrum Microphone 5

We know from theory that if these noises are independent then the added spectrum (figures 10 and 11) must be the same as the spectrum of the combined noise. We then compared the added spectrum of noises to that of the spectrum of the combined noise. And since the spectrums were comparably different, we concluded that these noises must be dependent.

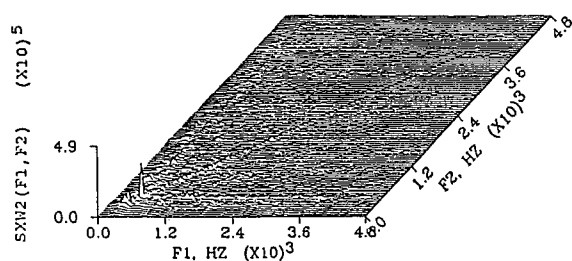


Figure 10. Added Spectrum Microphone 2

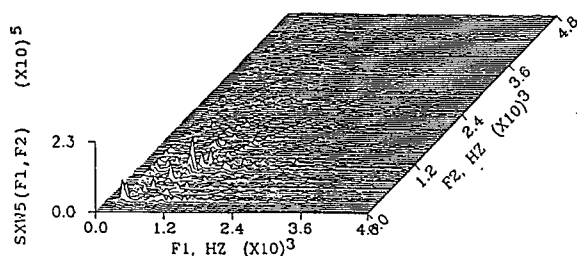


Figure 11. Added Spectrum Microphone 5

We note again that the combined noise is not necessarily PC, because it is the sum of two PC processes with incommensurate periods. It is probably a more general nonstationary process like CAR. According to blocked with respect to one period, we chose to block the data with respect to the period of the tail rotor noise. will naturally result in the estimated spectrum being skewed to some degree. However, the same amount of skewing occurs in the estimated spectrum for the added noise which is also blocked with respect to the tail rotor noise. Therefore, the estimated spectrum of the combined and added noise can be viewed equally with respect to theory.

### 5.2 Assumed Stationarity

The standard method for handling data taken from helicopter noise is to treat the data as if it were stationary. This results in a one dimensional spectrum along the diagonal. Therefore, the spectrum excludes all information concerning correlation of sinusoidal amplitudes at frequencies where  $\lambda \neq \theta$ . Although these neglected studied. An eample of a sound pressure spectrum, for the tail rotor produced using this method, is given in figures 12.

The graph produced contains 22 degrees of freedom and a frequency resolution of 49 Hz. There are noticeable peaks are the fundamental frequencies and subsequent harmonics. In the graph the pressure level drops off very rapidly. Notable peaks start to drop off after 3000 Hz. Due to the frequency range of audible sound (2 to 20 khz) we expect to find high pressure levels above 3000 Hz also. However, this is not evident with this technique. This leads us to study another technique which utilizes a random shift to reduce this deficiency.

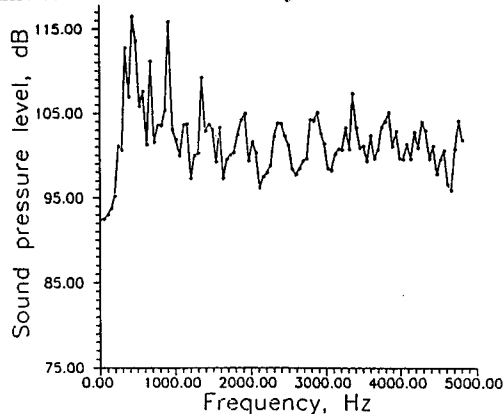


Figure 12. Tail Rotor Sound Pressure Level Microphone 3

### 5.3 Random Phase Shift

A PC process will become stationary by applying a random phase shift which is uniformly distributed over the period of the process. This stationary process can then be adequately analyzed by the standard method discussed above. This new stationary process which has been produced is no longer the original process. And thus, no longer contains information about the correlation of Fourier components at different frequencies. To implement the technique, we take a length of data and break it into blocks. Each block is of length necessary to obtain a desired frequency resolution plus the period of the underlying process. When the program is implemented, a random function call chooses a sample index uniformly distributed over the period of the process. Starting with the new sample values, frequency resolution. This is done for



each block of data. This technique results in wastage of some sample values from the start of the record, but does implement the random phase shift while maintaining the desired number of samples per block.

The sound pressure spectra produced from this shifted data can be compared to that of the original (unshifted) data. After viewing the graph in figure 13, it is apparent smoothed spectrum as if a moving average had been applied to the spectrum. There are additional peaks in the spectrum which correspond to the shaft frequency. Since the shaft frequency one fourth of the blade passage frequency, there could be three peaks between the harmonics of the spectrum for the tail rotor noise. This is evident in<sup>11</sup> in which a frequency resolution of 12 Hz is used. This resolution is fine enough to show the harmonics. However, our graph has a frequency resolution of 50 Hz which is not fine enough to completely show the shaft harmonics.

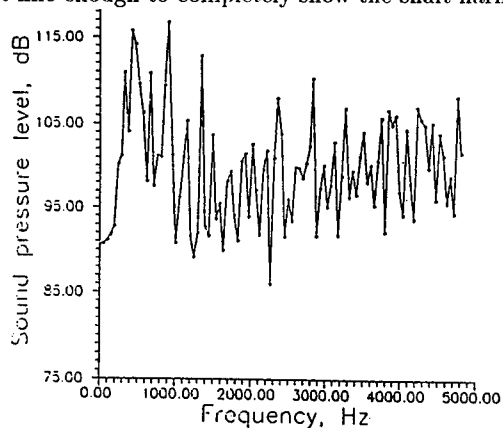


Figure 13. Tail Rotor Sound Pressure Level with Random Shift

This technique of shifting the data appears to be useful in harmonics of the transformed data. The output resulting from this method was also used in order to produce a two dimensional spectral estimate. The graph for stationary process from a PC process.

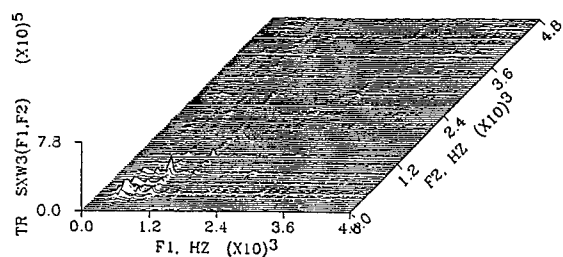


Figure 14. Tail Rotor Spectrum with Random Shift Microphone 3

### Conclusions

We discussed a method for producing a two dimensional spectral than considering nonstationary data as being stationary. We verified that applying a random phase shift to the data results in a more useful spectrum for viewing higher harmonics than when the data is method above appear to drop off to quickly. For a further study, we feel one must investigate the off diagonal values of the spectral estimate as to how their presence can be used to characterize the process itself.

### Acknowledgements

We wish to express our sincere gratitude to Dr. Jay C. Hardin, senior scientist at the NASA Langley Research Center, for his keen insight and sugestions throughout this work. This work has been supported by NASA grant NAG-1-768 andONR Grant N00014-89-J-1824.

## References

1. J. L. Doob, *Stochastic Processes* (John Wiley Sons, Inc., New York, 1953).
2. G. R. Dargahi-Noubary and A. G. Miamee, *Mathematical Geology* 25 (1993) 671-688.
3. W. A. Gardner, *Cyclostationarity in Communications and Signal Processing* (IEEE Press, New York, 1994).
4. E. G. Gladyshev, *Soviet Math. Dokl.* 2 (1961), 385-388.
5. J. C. Hardin and A. G. Miamee, *J. Sound and Vib.* 142 (1990) 191-202.
6. J. C. Hardin, A. Makagon, and A. G. Miamee, *Nonstationary Stochastic Processes and their Applications*, ed. A. G. Miamee ( World Scientific Publishing Co., New York, 1992) pp. 165-175.
7. A. G. Miamee and J. C. Hardin , *Sankhya Series A* 52(1990) 145-155.
8. A. G. Miamee and H. Salehi, *Multivariate Analysis*, ed. V. P. R. Krishnaiah (North-Holland, Amestrdam, 1980) pp. 167-179.
9. H. L. Hurd, *Proceedings of the 4-th Annual ASSP Workshop on Spectral Estimation and Modeling* ( IEEE Press, New York, 1988 ) p. 387.
10. H. L. Hurd, *J. Multivariate Analysis* 29 ( 1989) 53-67.
11. R. M. Martin, C. L. Burley, and J. W. Elliott,
12. A. Papoulis, *Probability, Random Variables, and Stochastic Processes* (McGraw-Hill, New York, 1984).

# MULTIFRACTAL CHARACTERIZATIONS OF INTERMITTENCY IN NONSTATIONARY GEOPHYSICAL SIGNALS AND FIELDS

## A MODEL-BASED PERSPECTIVE ON ERGODICITY ISSUES ILLUSTRATED WITH CLOUD DATA

ANTHONY B. DAVIS,<sup>†</sup> ALEXANDER L. MARSHAK,<sup>†</sup> WARREN J. WISCOMBE,  
AND ROBERT F. CAHALAN

NASA – Goddard Space Flight Center, Climate and Radiation Branch (Code 913),  
Greenbelt, MD 20771, USA.

E-mail: davis@climate.gsfc.nasa.gov

### ABSTRACT

There are many reasons for wanting to quantify spatio-temporal correlations in geophysical signals  $f(x)$  over a large range of scales  $r$ . Standard approaches use either the autocorrelation function  $\langle f(x+r)f(x) \rangle$  or the related 2nd-order structure function  $\langle [f(x+r)-f(x)]^2 \rangle$ , equivalently (Wiener-Khinchin theorem), the wavenumber spectrum  $E(k)$ , with  $k \approx 1/r$  as the scale parameter. These are all 2nd-order statistics however, and they do not discriminate well between fields with sometimes radically different spatial properties. For example, in seismic signals the background (possibly instrumental) noise can be modeled as white whereas the interesting events are more like Dirac  $\delta$ -functions: both components are  $\delta$ -correlated in the sense of  $\langle f(x+r)f(x) \rangle$  and have correspondingly flat wavenumber spectra  $E(k) \equiv \text{constant}$ . In another instance, temporal fluctuations of air temperature are Brownian motion-like, with  $\langle [f(x+r)-f(x)]^2 \rangle \propto r$  (hence  $E(k) \propto k^{-2}$ ), under quiescent meteorological conditions; unfortunately, the same spectrum is assigned to the occurrence of a quasi-discontinuity marking the passage of a front, as approximated by a Heaviside step function. The issue at hand is resolved by introducing the notion of “intermittency,” a concept borrowed from turbulence theory that describes the occurrence of bursts of intense events; statistically speaking, we are faced with the break-down of the prevailing Gaussian paradigm in data analysis. To characterize intermittency, some form of wavelet-type time/frequency (or position/scale) analysis is required. Multifractal approaches to position/scale analysis are particularly easy to exploit: they use higher-order moments as a simple way of sorting the continuum of weak, intermediate, and strong events, and we look for power-law regimes in the resulting scale-dependent statistical quantities at all orders. The two main categories of multifractal analysis,  $q$ th-order structure functions and singularity analysis, are surveyed and illustrated with both models and cloud-related data in 1D and 2D. We address in detail the sampling (or “ergodicity”) problems that arise as soon as Gaussian assumptions are relaxed and their relation to both stationarity and intermittency is discussed. Finally, we outline how multiscaling has helped to further the theory of cloud-radiation interaction, as applied to the forcing of the climate system and the remote sensing of cloud properties.

<sup>†</sup>Also: Science Systems and Applications, Inc. (SSAI), 5900 Princess Garden Parkway, Lanham, Md.

## 1. Introduction

### 1.1 Background

There is an increasing need in the geophysical community for statistical analysis of data. This need is traditionally met with techniques developed within entirely different areas of research. For example, first and second order statistics —means, variances and covariances— found their foremost applications in psychometrics. These well-known quantities are the parameters of the most general multivariate Gaussian distribution. Their counterparts in time-series analysis are the 2-point autocorrelation function and the wavenumber spectrum which proved to be powerful tools for solving engineering problems in communications and signal processing.

Over the decades, it became apparent that Gaussian —and otherwise “thin-tailed”— statistics were ill-suited to describe many random signals that reflect the variability of geophysical fields. Stretched exponentials, log-normal and even power-law distributions were introduced to describe seismic activity, rain-rates, atmospheric turbulence, and numerous other natural phenomena. Opening the Pandora’s box of non-Gaussian statistics with “long-” or “fat-tailed” distributions (Waymire and Gupta 1981) raises important questions about sampling. We will refer to these issues generically as “ergodicity” problems, a terminology that better reflects our model-based investigation. The basic question is: *Do space/time averages converge to well-defined ensemble counterparts with sample size and how fast? More formally put: How much of probability space do we need to explore to characterize a distribution?* For exactly Gaussian processes, the answer to these questions is more-or-less contained in the “3 $\sigma$ ” rule: events more than three standard deviations away from the mean are improbable at the level 0.001.<sup>1</sup> We will show further on examples of natural variability where the standard deviation itself is not even pinned down after many thousands of observations.

In our experience with geophysical data analysis, we have encountered at least two sources of ergodicity problems that generally appear compounded: “intermittency,” and “nonstationarity.” We are adopting vocabulary from time-series analysis here for simplicity but have either space or time in mind. [Rather than “(non)stationarity,” the technically correct usage in random field theory is statistical “(in)homogeneity.” However, in keep with cloud-modeling usage, we will reserve this last expression to designate (non-)constant fields which, in turn, are “(non-)trivial” from the times-series perspective.]

- It is natural to think that estimates of statistical properties of geophysical signals are independent of the instant when observations start and stop.<sup>2</sup> This is effectively a “stationarity” assumption, i.e., averages are statistically invariant under translations

<sup>1</sup>Every “Gaussian-type” or “thin-tailed” distribution has a similar rule; e.g., deviations in excess of  $\approx 4.9$   $\sigma$ ’s from the mean  $\mu$  are unlikely at the level  $10^{-3}$  for Laplace’s probability law:

$$\text{Prob}\{X \leq \xi < X+dX\} = \exp[-\sqrt{2}|X-\mu|/\sigma]dX/(\sqrt{2}\sigma).$$

<sup>2</sup>We assume here that all the observations belong to a well-defined class within which we can perform meaningful statistics. At least in atmospheric applications, this may impose external limits on the time of day, the season and the position on the globe.

in space or time. Observation time is irrelevant to systems in some kind of dynamical equilibrium. However, very long-range correlations can and do occur in geophysical signals because of the sheer size of the system and the coherent, long-lived structures generated by the large-scale forcing and the highly nonlinear character of the dynamics. In other words, stationarity is generally *not* a good assumption, at least at close range (small scales). The weaker assumption of nonstationarity with stationary increments is generally good enough at all scales (examples to follow). Only when the record is exceptionally long, do we observe a transition to stationarity *per se*. In summary, the questions to ask is '*for how long are the data correlated?*' or '*how long must we wait to isolate independent samples in a given datastream?*'

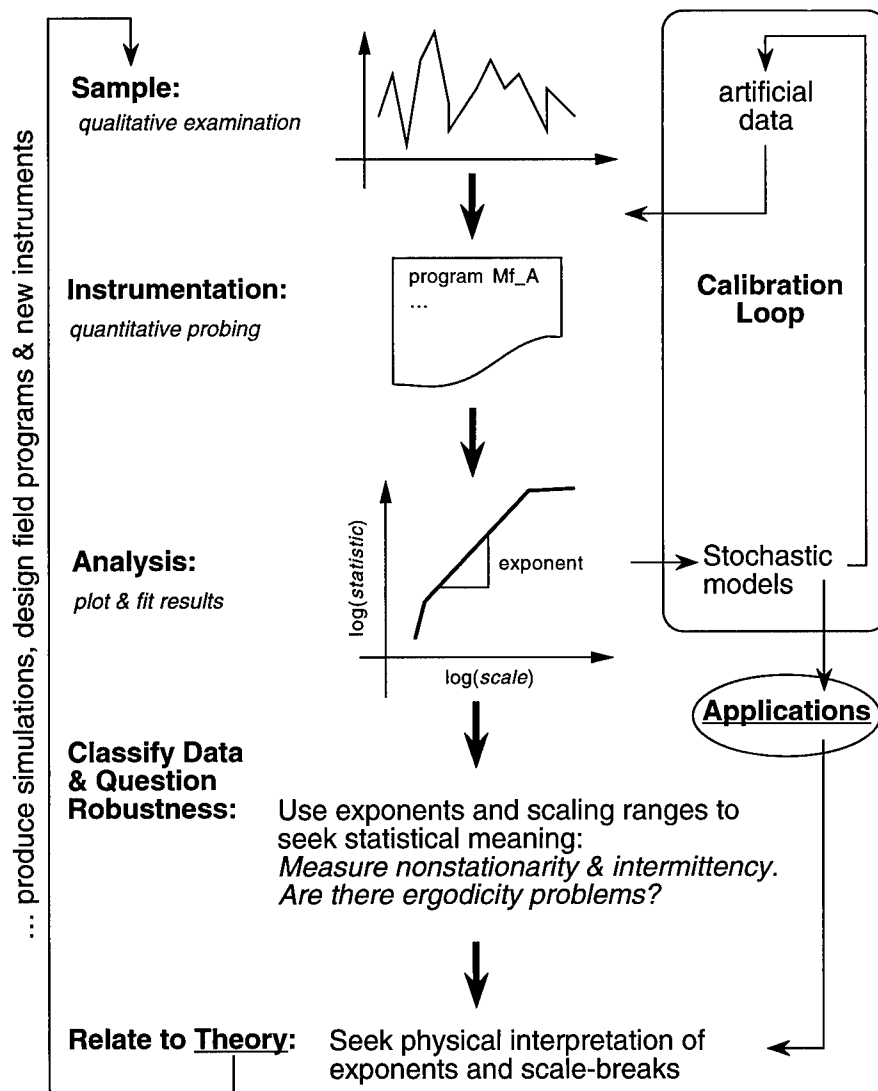
- Assuming that the nonstationarity has been tamed by focusing on the appropriate quantities to be averaged (e.g., increments), we can still be faced with "intermittency" problems. We borrow this concept from turbulence theory to describe the occurrence of sudden bursts of intense variability, very uncharacteristic of Gaussian processes —stationary or not. Following a now well-established tradition in turbulence studies (Parisi and Frisch 1985, Meneveau and Sreenivasan 1987a, etc.), we can use "multifractal" statistics to describe intermittency in natural signals. These are straightforward generalizations of the 2nd order statistics mentioned above where moments of all orders —within limits set by sampling considerations— are computed on a scale-by-scale basis and where the dependence on scale is parameterized by power laws. We thus define "scaling" regimes and associated families of exponents.

*A priori*, nonstationarity and intermittency are purely qualitative attributes when it comes to data analysis. Furthermore, some if not all of the positional information needed to make a statement about stationarity is lost due to the spatial averaging that produces the statistics in the first place. We will show that, in the framework of scale-invariant processes, nonstationarity and intermittency can both be not only detected (cf. §5.1 and §6.3) but precisely quantified as well (cf. §4.5).

Fractal (single-moment) and multifractal statistics were originally perceived as abstract and were criticized for having little bearing on the underlying physics. Serious efforts have been put forth to make multifractal concepts attractive to a broad range of geophysicists (e.g., Davis *et al.* 1994a), and their connections with wavelet analysis are now well-understood (Muzy *et al.* 1994). It is true that fractal concepts become mathematically precise in the small-scale limit ... which is generally unjustified on physical grounds. Nevertheless, "physical" fractals (with well-defined inner- and outer-scales) have proven to be very helpful models of reality in a broad range of applications. In fact, the limits of the scaling regimes themselves convey as interesting information on the system as do the exponents, if not more. We will survey our findings in this area with respect to cloud structure and cloud-radiation interaction.

## 1.2 Overview: The "Laboratory" Model for Geophysical Data Analysis

The paper is organized as follows —in the spirit of a report on a laboratory experiment (cf. Fig. 1). The goal of the "experiment" is to characterize the structure of



**Figure 1:** Flow-Chart for the "Laboratory" Model for Scale-by-Scale Statistical Analysis of Geophysical Data. There are two feed-back loops. One signifies that new theory makes new predictions to be verified with new data (from new instruments if necessary). The other represents the production of synthetic data with stochastic models to calibrate the statistical "instrumentation" with standard input. Stochastic modeling also feeds into the body of theory that explains the data. In the case of cloud-radiation theory, the numerical modeler can control the weather in his digital cloud system, the simplest models may be amenable to analytic methods.

marine stratocumulus for the purposes of radiative transfer computations. These are climatologically important cloud systems, and our main motivations to study them are (i) improved models for understanding the atmospheric radiation budget, hence present and future climate, and (ii) improved methods for retrieving cloud properties from remotely-sensed signals.

- Section 2 (“Sample Collection”): we present 1D and 2D data pertaining to marine stratocumulus and further motivate their study.
- Section 3 (“Instrumentation”) describes the basic tools of scale-by-scale statistical analysis: coarse-graining, autocorrelation, structure functions and spectral analysis.
- Section 4 (“Results”) establishes the relevance of power-law parameterizations for scale-dependent cloud statistics and defines notations for the associated exponents.
- Section 5 (“Semi-Empirical Criteria”) shows how we interpret certain scaling properties, with an emphasis on ergodicity issues: stationarity, intermittency and the onset of sampling problems are discussed.
- Section 6 (“Theoretical Considerations”): we rephrase our outlook on data analysis in terms of symmetry and broken symmetry; this impacts directly our understanding of cloud structure and how it transpires in satellite images. We also draw parallels between our methods and those of statistical physics.
- Section 7 is a summary.
- Appendix (“Calibration and Simulation”): a number of scale-invariant models are introduced and classified according to the criteria in Section 5. These algorithms for generating synthetic data have many applications, our present concerns being (1) assessment of the reliability of analysis procedures and (2) simulation of realistic clouds-in-a-computer that enable numerical radiation transport studies. In all cases, the statistical properties —namely, the exponents— are known *a priori*.

## 2. SAMPLE COLLECTION (Cloud Data as an Object of Study)

When designing an experiment, generally to test some hypothesis, the first questions to ask are ‘*What are we going to study?*’ and ‘*Why?*’. We can view data analysis as a straightforward experimental procedure where the object of study is the data itself (Fig. 1). We will assume it to be stored in a 1D or 2D array of real numbers residing in computer memory. In this section, we present geophysical data and present a rationale for an in-depth study of its statistical properties. The next questions are closely related: to each other: ‘*What properties are we interested in?*’ and ‘*What instruments will we use to probe our sample?*’. They are addressed in sections 3–4.

We present here data used in our specific research area: internal structure of marine stratocumulus (Sc) and its impact on radiation transport. Generally speaking, cloud-radiation interaction is a source of considerable uncertainty in the prediction of climate and climate change. Being both persistent and extended, marine Sc layers are responsible for a large portion of the Earth’s global albedo, hence the planet’s overall energetic equilibrium. A robust statistical characterization of marine Sc structure is therefore in order. In particular, this will allow us to develop statistically realistic cloud models



which can in turn be used to investigate radiation issues. An improved understanding of how physical cloud properties relate to their radiation fields has important spin-offs in the area of remote sensing. This is the only cost-effective way of monitoring cloud cover from synoptic to pixel scales (several kilometers or meters, depending on the device).

## 2.1 One-Dimensional in Situ Transects of Liquid Water Content from FIRE

### 2.1.1 Internal Cloud Structure Using Taylor's Frozen Turbulence Hypothesis

Clearly, there is no better way to study cloud structure than by direct probing. This calls for a fully instrumented aircraft and, because of the costs involved, few datasets of this type are available. Furthermore, cloud liquid water content (LWC) measurement is still an area of active research (Gerber *et al.* 1994). We present here transects of LWC in marine Sc that we will use to illustrate 1D data analysis in the remainder of the paper.

In Figs. 2a-e, we show representative samples of LWC vs. time from five flights (or flight legs) during FIRE<sup>1</sup> in June–July 1987 off the coast of southern California. Following a well-established practice in the turbulence literature (Taylor's frozen turbulence hypothesis), we perceive these time-series as 1D cuts through the spatially variable LWC field:

$$f_i(x_m), x_m = m\ell, m = 1, \dots, M_i \quad (i = 1, \dots, 5). \quad (1)$$

Table 1 shows the important parameters of the datasets, described in more detail by Davis *et al.* (1996a). In particular, they tentatively relate the down-spikes that characterize Figs. 2a,b to dynamical instabilities, and they question the reality of the strong “dip” in Fig. 2c. In the following analyses, events affected by this feature are ignored but Marshak *et al.* (1996) examine the consequences of not eliminating the spurious dip.

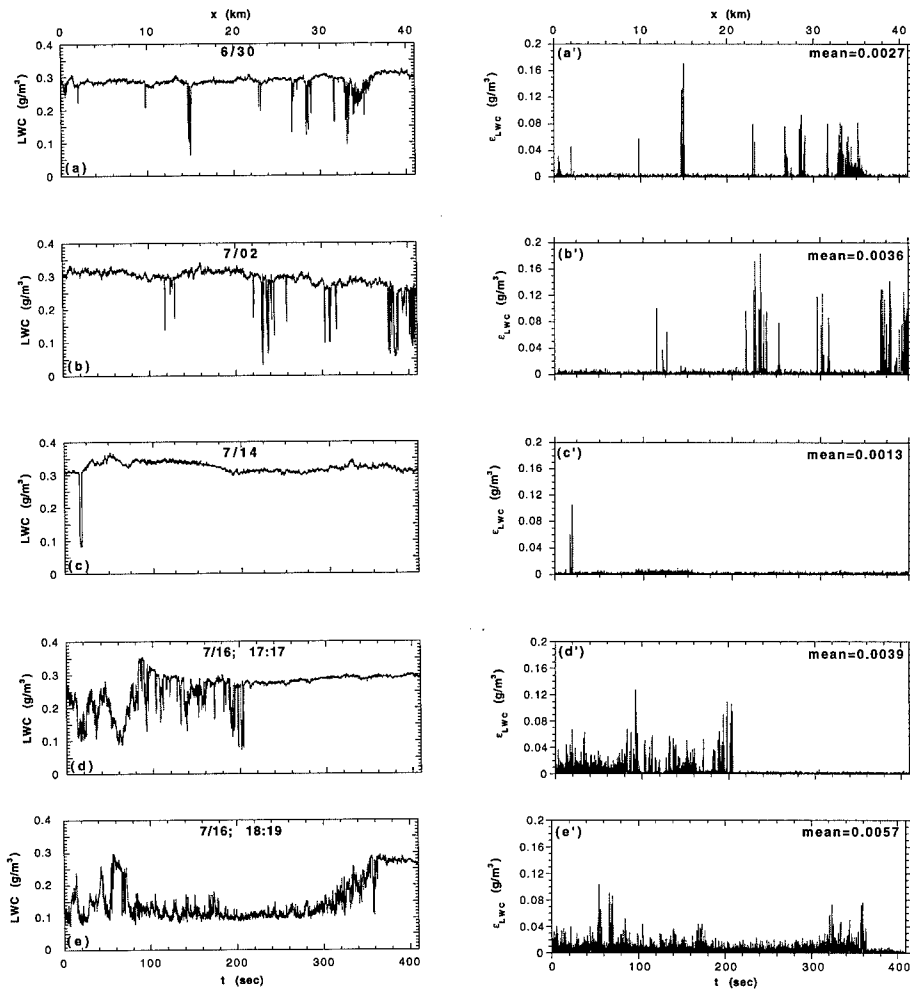
**Table 1:** FIRE Liquid Water Content (LWC) Database. The statistically relevant parameters of the various datasets are collated. They were obtained from an airborne platform during the FIRE 1987 stratocumulus experiment, off San Diego (Ca.). A nominal aircraft speed of 100 m/s was used to convert time to space, the sampling rate being 20 Hz.

Date (1987)	Time (GMT)	$M$ (points)	$\log_2 M$	Length $L$ (km)	Character	Comment
6/30	22:41	28672	14.81	143	spiky	downward
7/02	02:23	16384	14	82	spiky	downward
7/14	23:09	65536	16	328	smooth	suspicious dip
7/16	17:17	8192	13	41	mixed	spikes down
7/16	18:19	12020	13.55	60	mixed	spikes up

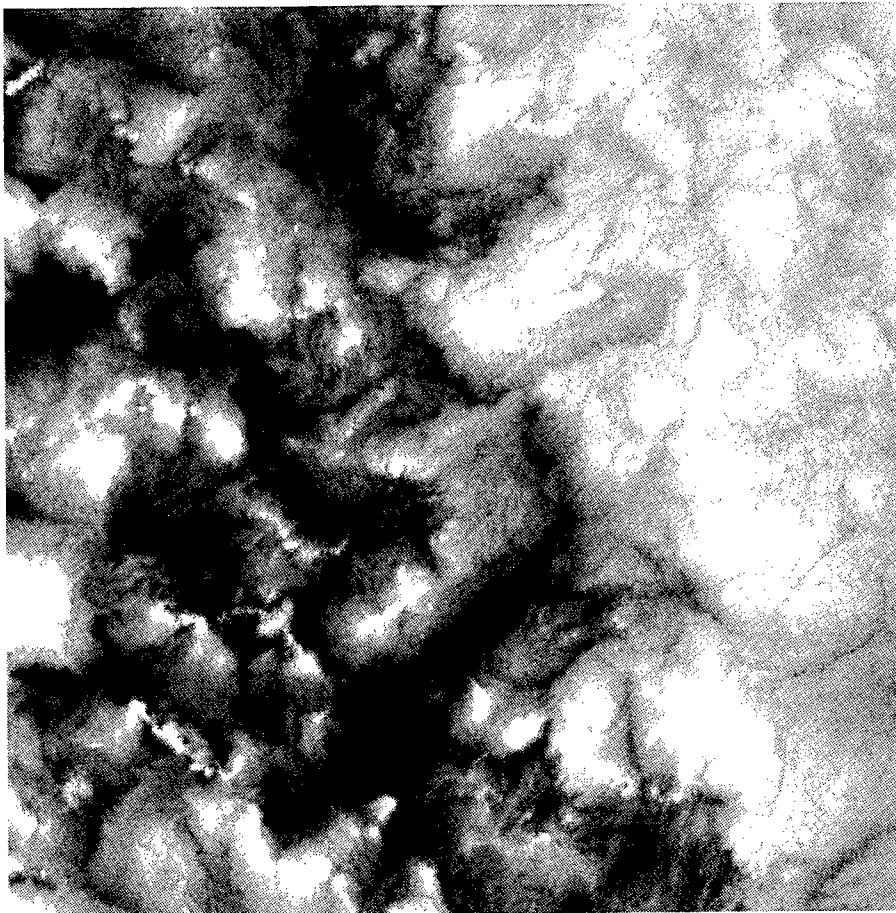
### 2.1.2 Visualizing Intermittency with Small-Scale Absolute Gradients

The most interesting (i.e., strong and somehow “organized”) features in Figs. 2a–e are the large and well-localized downwards deviations that occur intermittently but

<sup>1</sup> First ISCCP Regional Experiment (ISCCP = International Satellite Cloud Climatology Project).



**Figure 2:** Liquid Water Content (LWC) Transects from In Situ Airborne Probing of Marine Stratocumulus. (a–e) Representative portions of the five datasets described in Table 1; these are examples of nonstationary processes with stationary increments (Fig. 4 and §5.1) and multiscaling structure functions (Fig. 6 and §4.2). These data were collected in marine stratocumulus during FIRE in 1987 off the coast of southern California. (a'–e') Absolute next-neighbor differences for the data in panels (a–e); these are examples of intermittent stationary processes (Fig. 4 and §5.1) with *bone fide* multifractality (Table 2 and §5.3), as revealed by singularity analysis (Fig. 8 and §4.4).



**Figure 3:** *Radiance Field of Marine Stratocumulus.* Gray-scale rendering of a 4096×4096 portion of a Landsat image of a typical cloud deck off the coast of southern California. This scene was captured at visible wavelengths (channel 2 of the Thematic Mapper) on June 30 1987; so the climatological conditions are similar to those prevailing when the *in situ* LWC data in Fig. 2 were obtained.

nevertheless seem to cluster. The large jumps that characterize these events are highlighted by taking the absolute gradients of the data at some small (but presently unspecified) scale  $\eta$ . Adopting units where the sampling scale (grid constant)  $\ell = 1$  for simplicity, we have:

$$\epsilon_{ij}(\eta; x_m) = |f(x_{m+\eta}) - f_i(x_m)|, m = j, \dots, M_i - \eta + j \quad (j = 1, \dots, \eta; i = 1, \dots, 5). \quad (2)$$

The new data obtained from that in Figs. 2a-e with  $\eta = \ell = 1$  are presented in Figs. 2a'-e'.

In turbulence studies where  $f(x_m)$  is most often velocity,  $\eta$  is taken to be the "Kolmogorov" scale where dissipation forces start to dominate inertia. Furthermore, it is traditional in this context to take squares rather than absolute values in Eq. (2) since this yields a 1D cut through the physically important field that describes the local rate of kinetic energy dissipation (Meneveau and Sreenivasan 1987a).

## 2.2 A Two-Dimensional Radiance Field from LANDSAT Captured During FIRE

An attractive alternative to *in situ* probing of cloud structure is to use high-resolution satellite imagery. It is relatively inexpensive compared to outfitting and flying research aircraft, more comprehensive than aircraft probing or ground-based radiometry (being 2D), and more frequent in time.

Figure 3 is a gray-scale rendering of a large (4096×4096) section of a cloudy LANDSAT scene:

$$f(x_{m_1}, y_{m_2}), (x_{m_1}, y_{m_2}) = (m_1, m_2)\ell, m_1 = 1, \dots, 2^n, m_2 = 1, \dots, 2^n \quad (n = 12). \quad (3)$$

The signal  $f$  is digitized over 256 levels and is almost proportional to nadir-viewing radiance at satellite level ( $\approx 800$  km). LANDSAT's radiometer was not originally designed for such bright targets as clouds, so saturation (at  $f = 255$ ) occurs frequently, 17% in our case. In order to avoid spurious saturation effects, the up-coming statistical analyses use only the 2048×4096 leftmost pixels which are only 7% saturated.

## 3. INSTRUMENTATION (Scale-by-Scale Analysis Tools for 1D or 2D Datasets)

Pursuing our analogy between data analysis and laboratory work, we describe the first part of the experimental procedure. *What "measurements" are we going to do? What "instruments" are we going to use?* The object under scrutiny is cloud data stored in a large portion of computer (and possibly peripheral) memory. The instruments are programs that process this data; their output constitute statistical measurements. This new "data," residing in far less memory, describes partially the dataset. In essence, we are observing the statistical "behavior" of the data/subject with different devices and, in a sense we will define in §3.3, under different "experimental conditions."

All our instruments have two computational stages, performed in sequence (§3.1) or in parallel (§§3.2-4). First comes an analysis procedure that yields, in general, a quite large number of random variables by resampling and operating on the data. This is followed by a spatial/ensemble averaging. Consider, as an example, the computation of

1-point variance  $\langle f^2 \rangle - \langle f \rangle^2$ : it calls for (1) forming the 1st and 2nd powers of  $f(x)$  then (2) obtaining their averages over all the available data. In the following, we will consider exclusively 2- and more-point statistics that contain information about correlations (or “structure”) in the datastream. Since we always compute statistical properties at a specific scale  $r$ , we refer to these techniques collectively as “scale-by-scale” analyses.

### 3.1 Spectral Analysis

Correlations in random data can be studied via Fourier analysis, leading to energy spectrum estimation. This is a well-traveled approach to scale-by-scale analysis where the scale parameter is wavenumber  $k$ , related to the length scale  $r = 1/k$  in physical space.

Let  $\tilde{f}(k)$  be the  $d$ -dimensional Fourier transform of a field  $f(x)$  defined on  $[0, L]^d$ :

$$\tilde{f}(k) = \int_{[0, L]^d} f(x) \exp(2\pi i x k) d^d x, \quad (4a)$$

where the normalized wavevector  $kL$  scans  $\mathbb{Z}^2$ . In the discrete case, the above integrals become sums (with  $d^d x = \ell^d$ ):

$$\tilde{f}(k) = \ell^d \sum_x f(x) \exp(2\pi i x k). \quad (4b)$$

For  $d = 1$ ,  $k$  now goes from  $-k_N$ , excluded, to  $+k_N$ , included, where

$$k_N = \frac{1}{2\ell} \quad (5)$$

is the maximal (or “Nyquist”) wavenumber, by steps of  $\Delta k = 1/L_2$  with<sup>1</sup>

$$L_2 = M_2 \ell = 2^{\lfloor \log_2 M \rfloor} \ell, \quad (6)$$

$\lfloor \cdot \rfloor$  designating integer part. For  $d = 2$ , the subset  $(-k_N L_2, +k_N L_2] \otimes (-k_N L_2, +k_N L_2]$  of  $\mathbb{Z}^2$  is scanned by  $kL_2 = (k_x L_2, k_y L_2)$ . We note that, for  $f(x) \in \mathfrak{R}$ ,  $\tilde{f}(-k) = \tilde{f}(+k)^*$ .

In Fourier space, the spatial averaging step is replaced by a summation over phases, equivalently, over wavenumber sign. We compute the energy<sup>2</sup> spectrum in  $d = 1$  from

$$E(k) = \langle \|\tilde{f}(+k)\|^2 + \|\tilde{f}(-k)\|^2 \rangle \Delta k = \frac{2\delta_{0k}}{L} \langle \|\tilde{f}(k)\|^2 \rangle \quad (7a)$$

for  $kL \in \mathbb{N}$ , and where  $\delta_{0k} = 1$  for  $k = 0$ , 0 otherwise. In the discrete ( $\ell > 0$ ) case, the Fourier series is truncated, hence

<sup>1</sup>The most popular Fast Fourier Transform packages (e.g., Press *et al.* 1993) require  $M$  to be a power of 2. If this is not the case, the first and last  $M_2$  data points can be treated as two realizations in an ensemble average. Generally speaking, ensemble averaging over a number of datasets poses no special problem as long as they have the same  $\ell$  and  $L_2$ ; otherwise, common units for  $k$  and  $E(k)$  must be defined if different  $L_2$ 's occur and  $k$ -bins must be used if different  $\ell$ 's occur.

<sup>2</sup> $E(k)$  is called interchangeably “power” or “energy” or “wavenumber” or “frequency spectrum,” and sometimes “periodogram.”

$$E(k) = \frac{2\delta_0 k}{L_2} \langle \|\tilde{f}(k)\|^2 \rangle \quad (7b)$$

for  $kL_2 = 0, 1, \dots, L_2/(2\ell)$ . In  $d = 2$  cases, we assume statistical isotropy and sum  $\langle \|\tilde{f}(k)\|^2 \rangle$  over circles of radius  $k = |k| = \sqrt{k_x^2 + k_y^2}$  in Fourier space. For continuous spectra ( $\ell = 0$ ,  $L \rightarrow \infty$ , hence  $\Delta k \rightarrow 0$ ), we have

$$E(k) = \int_{|k|=k} \langle \|\tilde{f}(k)\|^2 \rangle d^2k = k \int_0^{2\pi} \langle \|\tilde{f}(k \cos \theta, k \sin \theta)\|^2 \rangle d\theta \quad (8a)$$

for  $k \geq 0$ . In the discrete ( $L < \infty$ ,  $\ell > 0$ ) case,  $kL_2$  goes from 0 to  $k_{\max}L_2 = [\sqrt{2}k_N L_2]$  and

$$E(k) = \frac{1}{L_2} \sum_{kL_2 \leq kL_2 < kL_2+1} \langle \|\tilde{f}(k)\|^2 \rangle \quad (8b)$$

for  $kL_2 = 0, 1, \dots, k_{\max}L_2$ .

It is sometimes advantageous to cumulate and average  $E(k)$  and  $k$  in octave-wide bins (i.e., by factors of 2 in  $k$ ) for  $k > 0$  (Davis *et al.* 1996a). There are precisely  $[\log_2 M] - 1$  bins when  $d = 1$ :  $\{kL_2 \in \mathbb{N}; 2^i \leq kL_2 < 2^{i+1} - 1\}$ , for  $i = 0, \dots, [\log_2 M] - 2$ ; this excludes the Nyquist frequency, the most aliased anyway (Press *et al.* 1993). In  $d = 2$  cases, one more bin can be populated by wavevectors with their modulus between  $k_N$  and  $k_{\max}$ : in all,  $\{kL_2 \in (-k_N L_2, k_N L_2]^2; 2^i \leq |k|L_2 < 2^{i+1} - 1\}$  for  $i = 0, \dots, [\log_2 M] - 1$ .

### 3.2 Two-Point Correlation Analyses at Order 2

There are two physical space counterparts of spectral analysis for correlation studies. First, one can form the products  $\delta f(x+r)\delta f(x)$  of the fluctuating part  $\delta f(x) = f(x) - \langle f \rangle$  of a 1D signal at two points and spatially (then, if necessary, ensemble-) average them to obtain the autocorrelation function:

$$\langle \delta f(x+r)\delta f(x) \rangle = \langle f(x+r)f(x) \rangle - \langle f \rangle^2. \quad (9a)$$

This well-known 2nd order 2-point statistic does not give us any new information since it is related to the energy spectrum in Eqs. (7–8) by the Wiener-Khinchin (W–K) relation:

$$\langle f(x+r)f(x) \rangle - \langle f \rangle^2 = \int_0^\infty \cos(2\pi rk) E(k) dk. \quad (9b)$$

At  $r = 0$ , the l.h.s. reduces to the 1-point variance, and the r.h.s. is the integral of  $E(k)$ ; so  $E(k)\Delta k$  is simply the part of the variance that comes from scales  $\approx 1/k$ .

One can also form “increments,”

$$\Delta f(r; x) = f(x+r) - f(x), \quad (10)$$

in 1D and compute the 2nd-order “structure function,”

$$\langle \Delta f(r; x)^2 \rangle = \langle [f(x+r) - f(x)]^2 \rangle, \quad (11a)$$

also known as a “semi-variogram” (Christakos 1992). Here again there is a W–K relation with the energy spectrum:

$$\langle \Delta f(r;x)^2 \rangle = 2 \int_0^\infty [1 - \cos(2\pi rk)] E(k) dk, \quad (11b)$$

as results from identities  $\langle \Delta f(r;x)^2 \rangle \equiv 2[\langle f^2 \rangle - \langle f(x+r)f(x) \rangle] \equiv 2[\langle f^2 \rangle - \langle f^2 \rangle - \langle \delta f(x+r)\delta f(x) \rangle]$ . Theoretically (i.e., when doing ensemble rather than spatial averages), these operations are only meaningful in “broad-sense” stationary situations where  $\langle \delta f(x+r)\delta f(x) \rangle$  depends only on  $r$ . However, Eq. (11b) generalizes to nonstationary signals with (broad-sense) stationary increments, namely, where  $\langle \Delta f(r;x)^2 \rangle$  is function of  $r$  alone.

In isotropic 2D situations,  $\delta f(x+ru)\delta f(x)$  or  $[f(x+ru) - f(x)]^2$  can be obtained by averaging over the allowable domain of  $x$  and the orientation of the unit vector  $u$ . This straightforward approach quickly becomes computationally intractable since it requires  $\sim N^2$  operations (where  $N = M^2$  is the total number of points). In contrast, FFT implementations of Eqs. (8b) or (11b) require only  $\sim N \ln N$  operations; in this case however, we interpret  $E(k)$  as the so-called “1D” spectrum obtained by dividing the r.h.s. of Eq. (8b) by  $\sum_{kL_2 \leq kL_2 < kL_2+1} 1$ , i.e., averaging rather than just summing  $\| \hat{f}(k) \|^2$ . Another approach (with only  $\sim N$  operations and generalizable to higher orders) is adopted in the remainder of this study: to treat rows and columns as an ensemble of 1D datasets.<sup>1</sup>

### 3.3 $q$ th-Order Structure Functions

How can one gain *new* information in the framework of 2-point statistics? Simply by looking at moments of order  $q \neq 2$ . The random variables of interest are then  $|\Delta f(r;x)|^q$  and averaging yields the  $q$ th-order structure function:<sup>2</sup>

$$\langle |\Delta f(r;x)|^q \rangle = \langle |f(x+r) - f(x)|^q \rangle. \quad (12)$$

Unfortunately, we lose the W–K connection and the computational efficiency of FFTs in 2 or more dimensions. However, the focus on increments is akin to a high-pass filtering. Therefore, at the cost of using the  $\sim N$  coefficients of a discrete wavelet decomposition of  $f(x)$  as surrogates for  $\Delta f(r;x)$ , the utilization of multiresolution analysis (Mallat 1989) will lead to efficient computational algorithms.

What insight do we gain by varying the parameter  $q$ ? Of all possible values that the increment  $\delta = |\Delta f(r;x)|$  can take, we can identify

- “typical” values that occur most frequently, being near the mode of the pdf  $p_r(\delta)$ ,
- “mean” values that dominate the average for  $q = 1$  (i.e., maximize  $\delta p_r(\delta)$ ),
- “r.m.s.” values that dominate the average for  $q = 2$  (i.e., maximize  $\delta^2 p_r(\delta)$ ).

There are also ever larger and rarer values that dominate higher order statistical moments:  $q = 3, 4$ , etc. So, increasing  $q$  amounts to looking at the more extreme values of  $|\Delta f(r;x)|$ .

<sup>1</sup>In theory, sampling is poor ( $\sim N$  out of  $\sim N^2$  possible events); but  $N = M_x \times M_y$  is already large in general.

<sup>2</sup>Validation of structure function computation and sensitivity studies with respect to amount of data can be performed with the help of the models presented in the Appendix, more specifically in §A.2 and §A.4.

This is akin to changing the experimental conditions (e.g., temperature  $T$ ) and observing changes in the state of a macroscopic physical system made of many macroscopic elements interacting with each other; for instance, a real gas or a ferromagnet. By lowering  $T$  the system can be forced into otherwise very unlikely configurations. The observable (macroscopic) changes can be very subtle or extremely dramatic, like when a phase transition occurs: e.g., a mole of  $\text{H}_2\text{O}$  molecules goes from vapor to liquid, to solid. Whatever the outcome, we learn more about the specificity of the system (data) by exploring as large a range of  $T$ 's ( $q$ 's) as possible.

### 3.4 Running Averages and the $q$ th-Order Moments of the Coarse-Grained Field

The high-pass filtering implicit when taking increments over various scales eliminates the large-scale mean value  $\langle f \rangle$  from the picture. What about the converse operation, computing local means (low-pass filter output) at various scales?

Consider for instance the running mean<sup>1</sup> of  $f(x)$  over  $[x, x+r]$ :

$$\mu_f(r; x) = \frac{1}{r} \int_x^{x+r} f(x') dx', \quad (13a)$$

for  $x = [0, L-r)$  and  $r > 0$ . In the 2D case, spatial averaging is over the square domain  $[x, x+r) \otimes [y, y+r)$  of area  $r^2$ . Consider also running variance of  $f(x)$  over  $[x, x+r]$ :

$$\sigma_f^2(r; x) = \frac{1}{r} \int_x^{x+r} f(x')^2 dx' - \mu_f(r; x)^2. \quad (13b)$$

In one application of the above running means and variances, we hold  $x$  constant and vary the scale  $r$  continuously until some kind of "convergence" is obtained (cf. discussion of Fig. 8a). Instead, we can think of  $r$  as a fixed parameter; then  $\mu_f(r; x)$  and  $\sigma_f(r; x)$ , obtained in Eqs. (13a, b), are random numbers. From this perspective, we can study their statistical properties by averaging over  $x$ , the position of the segment or square (cf. discussion of Fig. 8b), as well as seek the correlations<sup>2</sup> between  $\mu_f(r; x)$  and  $\sigma_f(r; x)$ .

By resampling  $x$  at intervals of length  $r$  in Eq. (13a), we obtain a "coarse-grained" version of the original field  $f(x)$ , with  $L/r$  pixels rather than  $L/\ell$ . This is of particular interest for the absolute gradients defined in Eq. (2). In the 1D continuum limit used in Eqs. (13a), we define

$$\epsilon(r; x) = \mu_\epsilon(r; x) = \frac{1}{r} \int_x^{x+r} \epsilon(x') dx', \quad r \geq 0, \quad (14a)$$

<sup>1</sup>In the language of continuous wavelet transforms, Eq. (13a) is the projection of the signal  $f(x)$  onto a functional space of "scaling functions." In this case, the scaling function equals  $1/r^d$  on a compact support of measure  $r^d$  and 0 elsewhere; Gaussian and otherwise variable "windows" have also been considered.

<sup>2</sup>This is the basis of the "spatial coherence" method developed by Coakley and Bretherton (1982) for recovering fractional cloudiness from satellite radiances.



and similarly in 2D. This translates to

$$\varepsilon(r;x) = \frac{1}{r} \sum_{x'=x}^{x+r-1} \varepsilon(1;x'), \quad r \geq 1, \quad (14b)$$

in the discrete case where the small-scale  $\varepsilon(1;x)$  field is defined on a grid with constant  $\ell = 1$ ; in other words, units of length where  $\eta = 1$  are employed in Eq. (2) when used to obtain  $\varepsilon(1;x)$  from  $f(x)$ . Generalization from 1D to 2D is again straightforward and, as for the wavelet-based surrogates for the  $\Delta f(r;x)$ 's mentioned in §3.2, efficient computation of  $\varepsilon(r;x)$  at appropriately selected  $x$ 's can be implemented in the framework of multiresolution analysis (cf. graphics by Davis *et al.* (1994a)).

Having obtained the non-negative random numbers in Eqs. (14a,b), we can take their  $q$ th powers and average the results over  $x$ , the positions of segments or squares, to obtain<sup>1</sup>  $\langle \varepsilon(r;x)^q \rangle$  with the same advantages in terms of experimental “temperature” control as discussed earlier for  $\langle |\Delta f(r;x)|^q \rangle$ .

#### 4. RESULTS (Power-Law Regimes and Scale-Breaks)

In this section, we examine the output of our laboratory “instruments.” These “measurements” are a new —and highly compressed— form of data: statistical quantities, always parameterized by scale. Typically, the scale parameter  $r$  or  $k$  spans a large range of values and, in general, the statistics (e.g.,  $E(k)$ ,  $\langle |f(x+r) - f(x)|^q \rangle$ , or  $\langle \varepsilon(r;x)^q \rangle$ ) do too. The usual way of visualizing quantities with large ranges is to use log-log plots and we naturally ask: *Are there significant ranges of scale where  $\log(\text{statistic})$  is linear in  $\log(r)$ ?* This is often the case and they are called “scaling” regimes. In such regimes the important parameter is the slope on the log-log plot, equivalently, the exponent in the associated power law. In this section, we introduce notations<sup>2</sup> for a number of exponents, present results for our cloud-related test data and discuss their most striking features.

##### 4.1 Scaling in Spectral Analysis

Figure 4a shows octave-binned energy spectra for the five 1D datasets in Table 1, partially illustrated in Figs. 2a–e. Log-log axes are used and the  $k$ -ranges are different due to the different lengths of the datasets ( $L_2$ 's). We see good scaling in all cases, at least for  $k_N/10^2 \lesssim k \lesssim k_N$  (last 7 octaves), with good agreement in the prefactors for 4 of the datasets. The dispersion at small  $k$  (large scales) reflects the visual diversity of Figs. 2a–e. The odd dataset is also the longest ( $L \approx 330$  km, about a half of all the FIRE LWC data), only an eighth of which is illustrated in Fig. 2c. Apart from the suspicious dip visible in that figure (but not incorporated in our analyses), this data looks very smooth. Spectrally, this translates in two ways: (1) the prefactor in the scaling regime is

<sup>1</sup>Validation of these computations and sensitivity studies with respect to amount of data can be performed with the help of models presented in the Appendix, specifically in §A.3.

<sup>2</sup>Unfortunately, there are several co-existing standards in the literature.

significantly smaller than for the four other datasets; (2) there is a specific scale  $1/k$  at which the Fourier modes stop increasing with scale and become constant. This is known as an “integral” (correlation) scale, denoted  $R$ ; we estimate  $R \approx 20\text{--}40$  km.

Figure 4b shows our results for octave-binned  $E(k)$  for the five absolute next-neighbor gradient fields associated with the LWC data analyzed in Fig. 4a. We notice that the behavior is somewhat more erratic than in Fig. 4a but still shows a reasonable correlation between  $\log E(k)$  and  $\log k$  for at least two decades in scale ( $\approx 7$  octaves).

Figure 4c shows the octave-binned estimates of the ensemble-average spectra for the LWC fields ( $f_i(x)$ ,  $i = 1, \dots, 5$ ) and for their absolute small-scale gradients  $\epsilon_i(\ell; x)$ , from Eq. (2) with  $\eta = \ell$  (in this special case, the second subscript is unnecessary). For both spectra, we see power-law behavior over no longer two but three decades (10 octaves) in scale. In other words, the realization-to-realization variability in  $E(k)$  at large scales has been damped by the ensemble-averaging. Unfortunately,<sup>1</sup> the break at  $R \approx 20\text{--}40$  km is not apparent in the average data because it starts at  $k = 1/\min\{L_2\}$  where  $\min\{L_2\} \approx 40$  km according to Table 1. However, we are confident that the transition from increasing to constant variance as scale increases ( $k$  decreases) will be consistently observed, whether at 20–40 km or more, as longer LWC datasets become available. This is because LWC is bounded on physical grounds: it is non-negative by definition with an upper bound which is a some small portion of the total amount of water vapor that the laws of thermodynamics allow a column of atmosphere to contain. This means that Fourier modes cannot be arbitrarily strong; a cut-off must occur when they reach the climatological mean LWC, otherwise fluctuations overwhelm the dc ( $k = 0$ ) component and the bounds will be exceeded.

The spectral scaling exponent  $\beta$  in

$$E(k) \sim k^{-\beta} \quad (15)$$

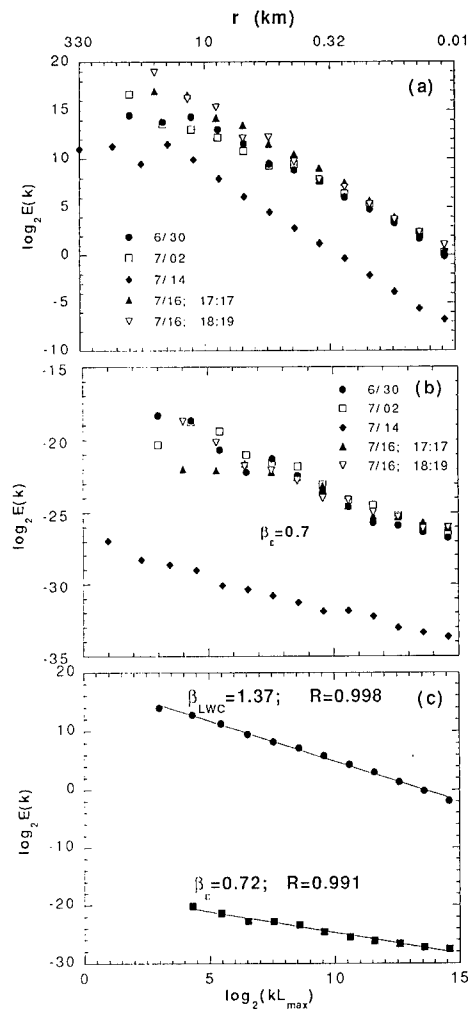
has quite different values for the two types of data:  $\approx 1.4$  for the LWC, and  $\approx 0.7$  for the associated  $\epsilon$ -fields; we discuss the statistical significance of this difference in §5.1.

Figure 5 shows  $E(k)$  vs.  $k$  in log-log axes for the l.h. half of the Landsat image in Fig. 3. In this case, there is so much averaging involved that octave-binning is not required to reduce the statistical noise. We see a clear-cut break at the integral scale: roughly constant Fourier amplitudes ( $\beta \approx 0$ ) for scales larger than  $R \approx 20$  km. There is also a break at  $\approx 0.2$  km, a scale we denote  $\eta_{rs}$ . Between  $1/R$  and  $1/\eta_{rs}$ ,  $E(k)$  follows a power law with  $\beta \approx 2.0$  and beyond  $1/\eta_{rs}$  a steeper law, with an exponent in excess of 3, the differentiability limit. In other words, small-scale variability is smoother than expected.<sup>2</sup>

Davis *et al.* (1996b) investigate the scale break at  $\eta_{rs}$  theoretically, relating it to a radiative smoothing phenomenon mediated by horizontal photon transport by multiple scattering. One outcome of this study is a simple expression for the “radiative smoothing” scale  $\eta_{rs}$  in terms of mean optical and geometrical cloud properties.

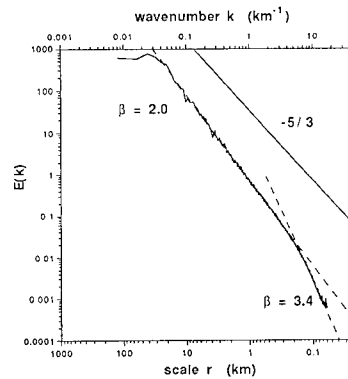
<sup>1</sup>Because very long records are needed, it is not always easy to estimate  $R$  in large geophysical systems.

<sup>2</sup>At the very smallest scales ( $\approx 60$  m, twice the pixel scale), another flattening occurs due to digitization.



**Figure 4:** Energy Spectra of 1D LWC data. (a) Five individual wavenumber spectra for the LWC datasets listed in Table 1, using octave-bins and log-log axes. (b) Same as (a) but for the absolute small-scale gradient fields. (c) Ensemble-average spectra from the data in panels (a) and (b), weighted by the overall lengths in Table 1.

**Figure 5:** Energy Spectrum of 2D Radiance data. Log-log plot of the wavenumber spectrum of the Landsat data in Fig. 3.



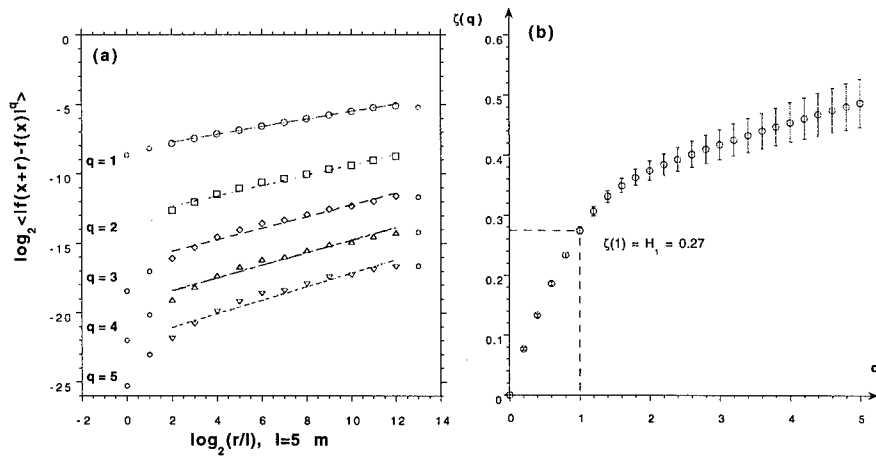


Figure 6: Structure Functions of 1D LWC data. (a) Scaling for  $q = 1(1)5$ . (b)  $\zeta(q)$  exponents.

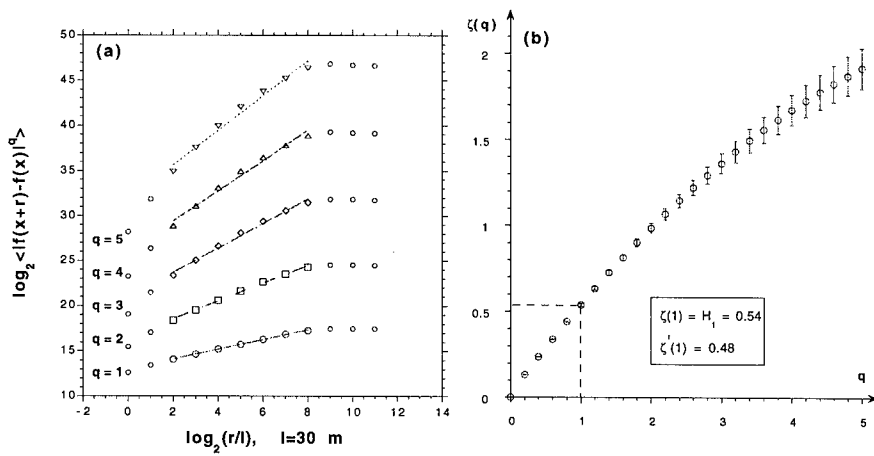


Figure 7: Structure Functions of 2D Radiance data. (a) Scaling for  $q = 1(1)5$ . (b)  $\zeta(q)$  exponents.

#### 4.2 Multifractal Analysis I, Higher-Order Structure Functions

Figure 6a shows the ensemble-average  $\log(|f(x+r)-f(x)|^q)$  vs.  $\log r$  for integer-valued  $q = 1, \dots, 5$  using the 1D LWC data as input. We note that by taking  $r = \ell$ ,  $r = 2\ell (= 1/k_N)$ ,  $r = 4\ell$ , etc., up to  $r = \min_{i=1, \dots, 5} \{L_{2i}\}$ , we obtain one more data point than with  $E(k=1/r)$ , namely, at the pixel scale  $\ell$ . This extra datum enables us to see some evidence of a break in the scaling at  $\approx 2\ell - 4\ell$  (10–20 m). At any rate, we see the break at the integral scale (i.e.,  $R \approx 20$  km) better than in the spectral data. Between these two limits, we have power laws for every  $q$ . Adopting the notation used in the turbulence literature, we have

$$\langle |f(x+r)-f(x)|^q \rangle \sim r^{\zeta(q)}. \quad (16)$$

Figure 6b shows  $\zeta(q)$  vs.  $q$  for  $0 \leq q \leq 5$ . Muzy *et al.* (1993) have developed an elegant method for estimating  $\zeta(q)$  directly from the modulus of continuous wavelet transforms.

Figure 7a shows the scaling of the  $q$ th-order structure functions for the l.h. half of the 2D data in Fig. 3, showing good scaling between  $\eta_{rs}$  and  $R$  and where both the scale-breaks, first characterized spectrally in Fig. 5, are apparent. (Here however, we would significantly underestimate the integral scale  $R \approx 20$  km by a factor of  $\approx 2$  due to the saturation at  $f(x) = 255$ .) Figure 7b shows the corresponding  $\zeta(q)$  function.

We note the concavity<sup>1</sup> of  $\zeta(q)$  and, since  $\zeta(0) = 0$  to ensure proper normalization, we can define a non-increasing hierarchy of exponents:

$$H(q) = \frac{\zeta(q)}{q}. \quad (17)$$

Data like ours yielding non-constant  $H(q)$ , hence a non-linear  $\zeta(q)$ , is “multifractal” in the sense of Parisi and Frisch (1985); otherwise (i.e., when  $\zeta(q) \propto q$ ), it is “monofractal.”

Two values of  $q$  are of particular interest. At  $q = 2$ , we retrieve the scaling for the 2nd order structure function; for power law statistics, the W–K relation in Eq. (11b) yields (e.g. Monin and Yaglom 1975, p. 92)

$$\beta = 2H(2)+1 = \zeta(2)+1. \quad (18)$$

Incidentally, relation (18) is well verified numerically by our data, especially if octave-bins are used in the spectral analysis (Davis *et al.* 1996a).<sup>2</sup>

At  $q = 1$ , we can retrieve the fractal dimension  $D_g$  of the graph of the data, viewed as a set  $g = \{x, f(x)\}$  embedded in  $d+1$  dimensions (e.g. Falconer 1990):

$$D_g = (d+1)-H(1) = (d+1)-\zeta(1). \quad (19)$$

<sup>1</sup>This property follows from characteristic function theory (cf. Feller 1971). The Fourier transform of the pdf of a real random variable  $\xi$  is its characteristic function  $\phi(t) = \langle \exp(it\xi) \rangle$ . Non-negativity of the pdf implies that  $\phi(iu)$  is real and non-increasing for real  $u$ . By the same token,  $\ln\phi(t)$ , the cumulant-generating function of  $\xi$ , is convex for purely imaginary  $t$ . Setting  $\xi = \ln|f(x+r)-f(x)|$  for fixed  $r$  and  $t = iq \ln r$ , we see from Eq. (16) that  $-\zeta(q)$  is proportional to a cumulant-generating function, hence the concavity of  $\zeta(q)$ .

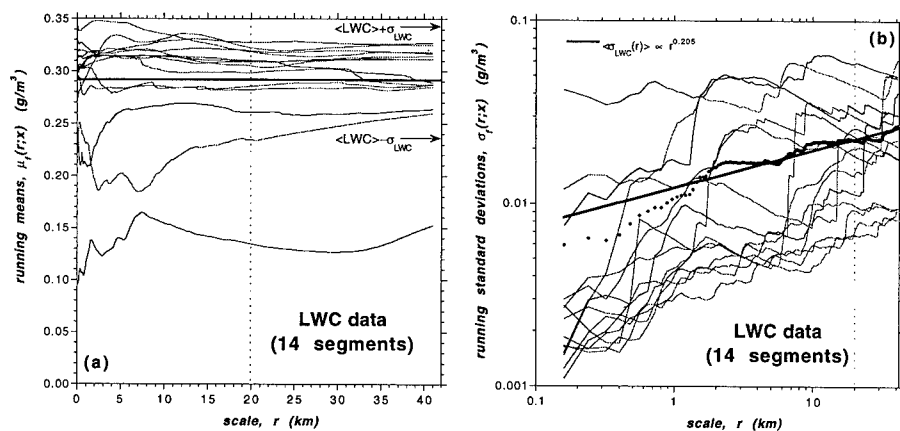
<sup>2</sup>This constitutes an internal calibration of the instrumentation in our data analysis lab; see Appendix for external calibration procedures using standard input.

We can therefore equate  $H(1)$ , the “mean Hölder exponent,” to the codimension of the graph of the data. This quantity has natural bounds. If  $H(1) = 0$ , then  $D_g = d+1$ : the graph fills a finite portion of  $\mathcal{R}^{d+1}$ . If  $H(1) = 1$ , then  $D_g = d$ : the graph is as smooth as the support  $\mathcal{R}^d$  of  $x$ , namely, a Euclidian line or plane.

### 4.3 Running Means and Variances

Figure 8a demonstrates the practical importance of the integral scale  $R$ —a 2-point statistical construct—for a 1-point statistic which is *a priori* simpler. We have plotted  $\mu_f(r;x)$  vs.  $r$  for 14 different non-overlapping 40 km sections from the LWC database. Running means begin to stabilize only at  $r \approx 20$  km, some to values quite far from the ensemble-mean of  $0.29 \text{ g/cm}^3$  (beyond one ensemble- $\sigma$ ). This tells us two things. First, we need a stretch of at least one or two integral correlation lengths before we can even talk about a mean, even locally. Second, this data is highly non-ergodic: no physically attainable length of LWC data seems to be enough to reach a “climatological” average.<sup>1</sup>

Figure 8b shows  $\sigma_f(r;x)$  vs.  $r$  for the same 14 sections, this time in log-log axes. The striking feature is the jumpiness of running variance due to localized events, clearly the intermittently distributed clusters of down-spikes in Figs. 2a,b,d,e; in one case, a factor of  $\approx 2$  is gained even after accumulating for  $\approx 30$  km. Such jumps are very unlikely in a process obeying Gaussian statistics. The r.m.s. ensemble-average,  $\langle \sigma_f^2(r;x) \rangle^{1/2}$ , is also plotted (bold dots): for  $r$  between 0.15 and 20 km, it follows a power law in  $r^{0.2}$ . The exponent is numerically identical to  $\zeta(2)/2 = (\beta-1)/2m$ , not a coincidence:  $\langle \sigma_f^2(r;x) \rangle$  should be of the same order of magnitude as  $\langle [f(x+r)-f(x)]^2 \rangle$ . Here again, no convergence to the estimated ensemble-mean  $\sigma$  ( $0.05 \text{ g/cm}^3$ ) is in sight, even at  $r = 40 \text{ km} \geq R$ .



**Figure 8:** Running Averages for 1D LWC Data. (a)  $\mu_f(r;x)$  vs.  $r$  for  $0 \leq r \leq 40$  km and 15 different locations separated by 40 km or more. (b) Same as (a) but for  $\sigma_f(r;x)$ .

<sup>1</sup>We are in no position to claim that  $0.29 \text{ g/cm}^3$  is climatologically relevant, not even for marine Sc in summer off of southern California (where FIRE was conducted in June-July 1987). Moreover, even for a given cloud-type, location, and season, we have 5 means (also true of characteristics described below).

#### 4.4 Multifractal Analysis 2, Singular Measures

Figure 9a shows the spatial/ensemble-average<sup>1</sup>  $\log_2 \langle \varepsilon(r; x)^q \rangle$  vs.  $\log_2 r$  for integer-valued  $q = 1, \dots, 5$  and  $\log_2(r/\eta) = 0, \dots, [\min_i = 1, \dots, 5 \log_2(L_i/\eta)]$  using the LWC data in Eq. (2) with  $\eta \approx 4\ell$  (20 m) and then Eq. (15). For  $r \geq \eta$ , we have power laws for all  $q$ 's; adopting the notation of Schertzer and Lovejoy (1987), we posit:

$$\langle \varepsilon(r; x)^q \rangle \sim r^{-K(q)} \quad (20)$$

for the  $q$ th-order moments of the coarse-grained measures. In Fig. 9b we show  $K(q)$  versus  $q$  for  $0 \leq q \leq 5$ .

An often used alternative to the  $d$ -dimensional *coarse-grained* measure  $\varepsilon(r; x)$  is the *total* measure in the interval  $[x, x+r)$  or domain  $[x, x+r) \otimes [x, x+r)$ :

$$p(r; x) = r^d \varepsilon(r; x). \quad (21a)$$

In this case, sums—not averages—of the  $q$ th power of  $p(r; x)$  over the  $(L_2/r)^d$  disjoint intervals are used:

$$Z_r(q) = \sum_x \langle p(r; x)^q \rangle \sim r^{\tau(q)}, \quad (21b)$$

where  $\langle \cdot \rangle$  now designates an (optional) ensemble-average over different realizations of  $\varepsilon$ . By separating spatial and ensemble averages in (20)  $(L_2/r)^d \langle \varepsilon(r; x)^q \rangle = r^{-dq} \sum_x \langle p(r; x)^q \rangle$ . Hence, from Eqs. (20) and (21b), we find

$$\tau(q) = (q-1)d - K(q). \quad (22)$$

Methods for estimating  $\tau(q)$  using continuous wavelet transforms have been developed by Arnéodo *et al.* (1988).

Two  $K(q)$  values are predetermined. Normalization requires  $K(0) = 0$  ( $\tau(0) = -d$ ). Only at  $q = 1$  can we permute the spatial averaging inside the interval  $[x, x+r)$  and the spatial average over the various  $x$ 's. In the convention where  $x$  is sampled every  $r$ , there is no difference in the outcome as  $r$  is varied:  $K(1) = 0$  ( $\tau(1) = 0$ ).

We note the convexity of  $K(q)$  in Fig. 9b, the associated  $\tau(q)$  being concave. This remarkable property is traceable to the same probabilistic cause as for  $\zeta(q)$ 's concavity. Here again this can be used to define the non-increasing hierarchy of exponents:

$$D(q) = \frac{\tau(q)}{q-1} = d - \frac{K(q)}{q-1}, \quad (23)$$

the generalized dimensions introduced originally by Grassberger (1983) and Henchel and Procaccia (1983) for the characterization of strange attractors in deterministic chaos theory.<sup>2</sup> Some dimensions are noteworthy; in standard terminology, we have:

<sup>1</sup>It is customary to resample  $x$  every  $r$  (from 0 to  $L_2-r$ ) in the spatial part of the averaging procedure; this guarantees that no data is used more than once for a given statistic.

<sup>2</sup>In that context,  $p(r; x)$  is the number of points sampled in the phase space of the dynamical system.

- $D(0^+) = d + K(0^+) = \tau(0^+)$ , dimension of the support;<sup>1</sup>
- $D(1) = d - K'(1) = \tau'(1)$ , information dimension;
- $D(2) = d - K(2) = \tau(2)$ , correlation dimension.

At  $q = 2$ , the scaling of  $\langle \varepsilon(x+r)\varepsilon(x) \rangle$  can be related to that of  $\langle \varepsilon(r;x)^2 \rangle \sim r^{-K(2)}$ , at least for multiplicative cascade models (see Appendix, §A.3.2). From the power-law translation of the Fourier duality in Eq. (9b), we have

$$\beta_\varepsilon = 1 - K(2) = 1 - [d - \tau(2)] = (1 - d) + D(2). \quad (24)$$

A fourth  $q$ -value was attracted considerable attention:

- $D(q_D) = 0$  or  $\tau(q_D) = 0$  or  $K(q_D) = (q_D - 1)d$  define the “critical” moment  $q_D$ .

For  $q \geq q_D$ , the moments of  $\varepsilon(r;x)$  are divergent (Mandelbrot 1974, Kahane and Peyrière 1976, Schertzer and Lovejoy 1987, Gupta and Waymire 1993, and others).

Data like ours, yielding a nonlinear  $K(q)$  hence a non-constant  $D(q)$ , are “multifractal” in the sense of Halsey *et al.* (1986), Meneveau and Sreenivasan (1987a,b), Schertzer and Lovejoy (1987), Evertsz and Mandelbrot (1992), and others. If  $D(q) \equiv \text{constant}$ ,  $K(q)$  and  $\tau(q)$  are proportional to  $q-1$ , and the data are said to be “monofractal.”

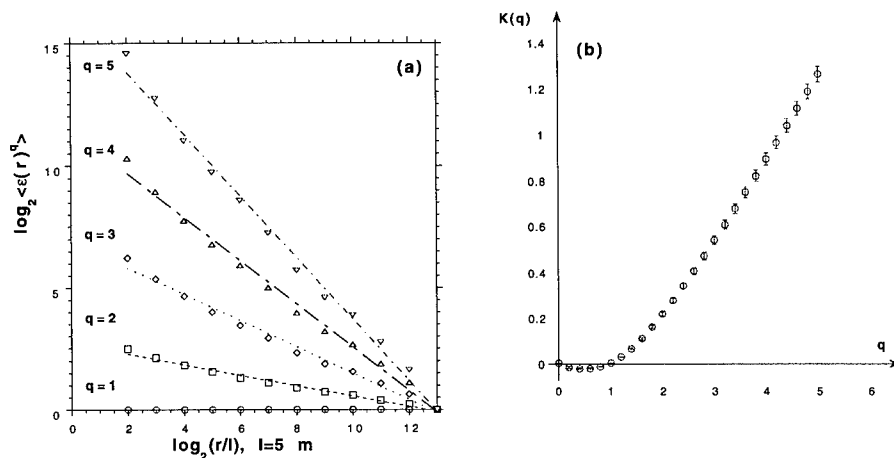


Figure 9: Singular Measures for 1D LWC Data. (a) Scaling for  $q = 1(1)5$ . (b)  $K(q)$  exponents.

#### 4.5 Bi-Fractal Analysis

Is it strictly necessary to have two kinds of multifractal analyses? In other words: Is there a general connection between  $\zeta(q)$  and  $K(q)$ ? This is an open question discussed at length in the specialized literature (Schertzer and Lovejoy 1987, Meneveau and Sreenivasan 1991, Sreenivasan 1991, Frisch 1991, Davis *et al.* 1993, Vainshtein *et al.* 1994), the consensus being that at least one extra exponent is needed to go from the

<sup>1</sup> $q = 0^+$  means that  $\varepsilon(r;x) > 0$  gives 1 and  $\varepsilon(r;x) = 0$  gives 0. Note that  $K(0^+) \leq 0$ , due to convexity.



currently fashionable  $K(q)$ -based approach to the more traditional one based on structure functions and  $\zeta(q)$ . Let this exponent be

$$H_1 = H(1) = \zeta(1); \quad (25a)$$

Bertozzi and Chhabra (1994) show that  $\kappa = 1 - H_1$  is the "cancellation" exponent recently introduced by Ott *et al.* (1992) to measure the rate at which  $f(x)$  goes from an up-trend to a down-trend. We prefer to view the (mean) Hölder exponent  $H_1$  as an index of nonstationarity: the degree of nonstationarity increases with  $H_1$ , ranging from 0 (stationarity) to 1 (differentiability).<sup>1</sup>

Another question of practical importance is: *How many  $q$ 's do we actually need in any given application?* It is impossible to answer this question on general grounds. However, we consider it important in any application to get at least a first-order estimate of the degree of intermittency in the system. Using the small-scale absolute gradients is a convenient way of doing this and a 1st-order index of intermittency is the information codimension

$$C_1 = d - D(1) = d - \tau'(1) = K'(1). \quad (25b)$$

At  $C_1 = 0$ , there is no intermittency. At  $C_1 = d$ , intermittency is maximal, corresponding to a situation where all the measure is concentrated onto a finite number of points (cf. Dirac measures in the Appendix).

Relations of the type

$$\zeta(q) = q/a - K(q/b) \quad (26)$$

have been proposed where  $a$  and  $b$  are constants. For instance,  $a = b = 3$  for turbulent signals (e.g., Sreenivasan 1991) and  $a \leq b$  for a model of Schertzer and Lovejoy's (1987) described in the Appendix. The generality of this  $\zeta(q) \leftrightarrow K(q)$  connection remains an open question.<sup>2</sup> If however Eq. (26) is either true in general, accepted as an approximation for low enough  $q$ 's, or used as a definition of  $K(\cdot)$  after setting  $b$ , then one can derive both  $H_1$  and  $C_1$  from structure functions alone, without resorting to measures based on gradients. Indeed, Eqs. (25–26) yield

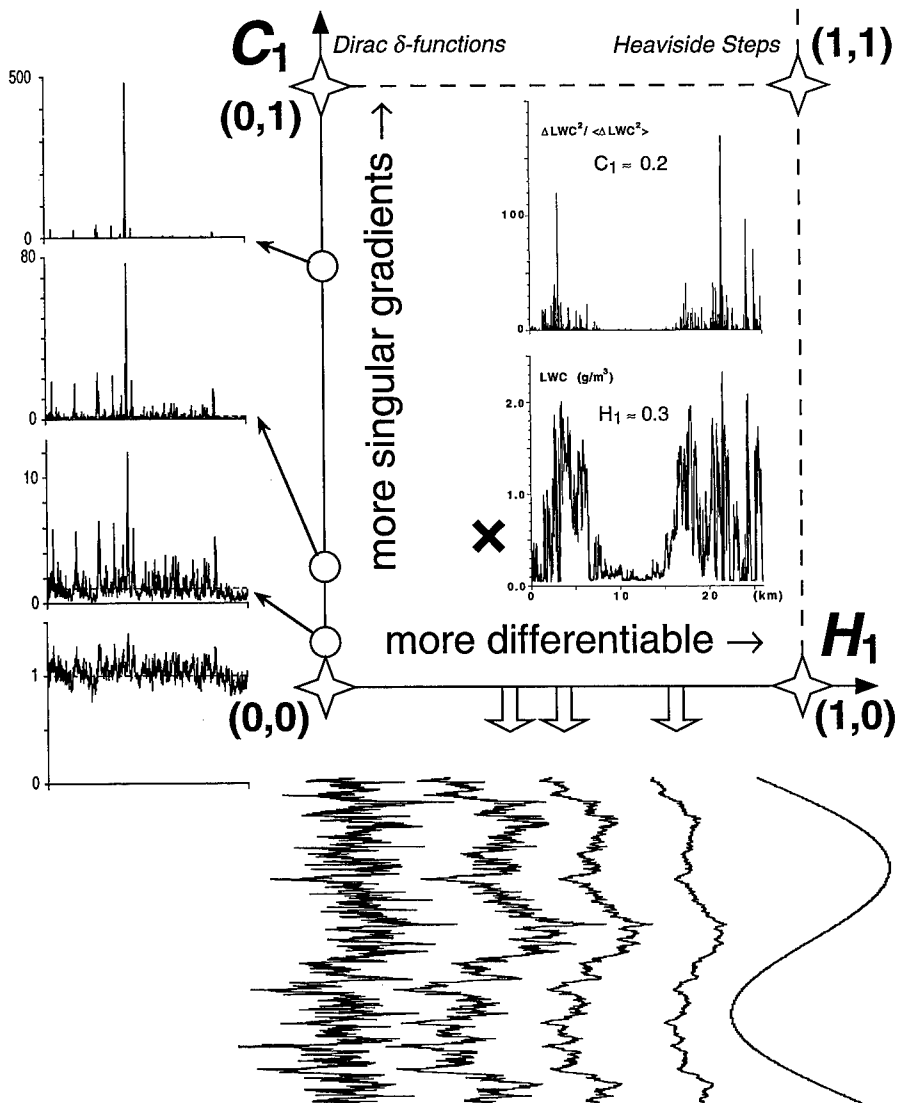
$$\begin{cases} H_1 = \zeta(1) = 1/a - K(1/b) \\ C_1 = K'(1) = b/a - b\zeta'(b) = \zeta(b) - b\zeta'(b) \end{cases} \xrightarrow{b=1} \begin{cases} H_1 = \zeta(1) = 1/a \\ C_1 = K'(1) = \zeta(1) - \zeta'(1) \end{cases} \quad (27)$$

Notice that if  $\zeta(q) \propto q$ , then  $K(q) \equiv 0$ , hence  $C_1 = 0$ ; so  $C_1$  is a measure of the curvature in  $\zeta(q)$ . As an example, we obtain  $C_1 \approx 0.06$  in this manner for the 2D Landsat data using  $b = 1$  for simplicity (cf. Fig. 7).

"Bi-fractal" analysis is the minimal form of multifractal analysis based solely on  $H_1$  and  $C_1$ . [Davis and Marshak (1996) discuss, compare, and relate other choices used in the literature.] Figure 10 shows a schematic  $(H_1, C_1)$  plane for  $d = 1$ . In our experience with marine Sc, bi-fractal analysis has proven very useful, leading in particular to new

<sup>1</sup>In §5.1 we will show that processes with  $\beta < 1$  are stationary; those with  $1 < \beta = 2H(2) + 1 < 3$  are nonstationary with stationary increments; those with  $\beta > 3$  are almost everywhere differentiable and have nonstationary increments. Since  $H(2) \leq H(1) = H_1$ , we have  $0 < H_1 < 1$ .

<sup>2</sup>A methodology for empirical verification is suggested by Davis *et al.* (1993).



**Figure 10: The Bi-Fractal Plane.** This schematic shows how the vertical coordinate reflects the intermittency of the squared small-scale gradients (lower inset) whereas the horizontal coordinate characterizes the nonstationarity of the primary signal (upper inset). Synthetic data with increasing  $C_1$  and  $H_1$  grace the axes. We see that, contrary to spectral analysis, bi-fractal analysis distinguishes Brownian motion,  $(H_1, C_1) = (1/2, 0)$ , from Heaviside steps,  $(H_1, C_1) = (1, 1)$ , which are both nonstationary ( $H_1 > 0$ ). The same is true for the stationary ( $H_1 = 0$ ) gradients of these two theoretical cases: white noise has  $(H_1, C_1) = (0, 0)$ , and a single spike, modeled by a  $\delta$ -function, has  $(H_1, C_1) = (0, 1)$ .

insight into cloud dynamical processes. Table 2 gives  $(H_1, C_1)$  coordinates for the data presented in section 2.

**Table 2:**  $(H_1, C_1)$  values for FIRE Landsat radiances and LWC, ASTEX and SOCEX LWC, and ARM LW path data.

Database (dataset)	$H_1$	$C_1$	$r_{\min} / r_{\max}$	Reference
FIRE Landsat radiances (6/30)	0.54	0.06	150 m / 10 km	(this publication)
FIRE LWC (6/30, 22:41Z)	0.29	0.14	20 m / 20 km	Marshak <i>et al.</i> (1996)
FIRE LWC (7/02, 02:23Z)	0.22	0.15	" "	" " "
FIRE LWC (7/14, 23:09Z)	0.34	0.03	" "	" " "
FIRE LWC (7/16, 17:17Z)	0.31	0.08	" "	" " "
FIRE LWC (7/16, 18:19Z)	0.34	0.07	" "	" " "
FIRE LWC 1987 ensemble	0.28	0.10	" "	" " "
ASTEX LWC 1992 ensemble	0.29	0.08	60 m / 60 km	Davis <i>et al.</i> (1994a)
SOCEX LWC 1994 ensemble	0.28	0.09	5 m / 5 km	(in preparation)
ARM LW path ensemble	0.37	0.08	1 min / 8 hr	Wiscombe <i>et al.</i> (1994)

The scatter, most notable in the intermittency index, reflects the diversity in Figs. 2a–e and argues for a non-ergodic model for this data. FIRE was conducted off the coast of Southern California, near San Diego. We have added to Table 2  $(H_1, C_1)$  entries for ensemble averages from analyses of LWC in marine Sc for two other field programs: the Atlantic Stratocumulus Transition EXperiment (ASTEX) that focused on a more complex situation with transitions to cumulus regimes, and the Southern Ocean Coupled EXperiment (SOCEX) that was conducted off the coast of Australia. The proximity of the ensemble average  $(H_1, C_1)$  points for the three different local climates argues for a degree of universality in the dynamics that determine the internal structure of marine Sc. The  $H_1$ 's for LWC cluster near 1/3, the value that characterizes turbulent fields: velocity (Kolmogorov 1941), temperature, and passive admixtures (Obukhov 1949, Corrsin 1951). The  $C_1$ 's for LWC cluster near 0.1, precisely in the range observed for turbulent signals.<sup>1,2</sup> Thus, although H<sub>2</sub>O (in all its phases) is far from being a dynamically passive constituent of the atmosphere, it can be perceived as advected by the turbulent wind field to a first approximation.

Finally, we have appended to the above LWC statistics an ensemble-average  $(H_1, C_1)$  for liquid water path (column integrated LWC) measured at the ARM site in Lamont (Ok.) for arbitrary cloud cover, as opposed to Sc only. We find roughly the same  $C_1$  but a somewhat larger  $H_1$  than for the three LWC observations. This is not surprising: the vertical integration that relates LWC to LW path will produce a smoother signal. LW path is more directly relevant than LWC to the radiative properties of the cloud, as

<sup>1</sup>  $C_1$  values can be obtained from "intermittency parameters,"  $K(2)$ , gleaned in the literature. To go from the characterization at  $q = 2$  to  $q = 1$ , two extreme hypotheses are log-normality,  $K(q) = C_1 q(q-1)$ , and monoscaling,  $K(q) = C_1(q-1)$ . Allowing for this uncertainty ( $C_1/K(2) \approx 1-2$ ), so  $C_1$  falls in 0.2–0.3.

<sup>2</sup> In turbulence studies, the dissipation rate field is obtained by squaring the velocity gradients at the Kolmogorov scale, leading to  $K_2(q)$ , rather than taking their absolute values, leading to  $K_1(q)$ , a difference in methodology easily accounted for: Lavallée *et al.* (1993) show that  $K_1(q) = K_2(q/2) - qK_2(1/2)$  which puts  $C_1 = K_1'(1)$  in the range 0.07–0.15.

observed in the Landsat image. The  $H_1$  for marine Sc radiance fields in visible channels is even smoother than for LW path because the escaping radiation fields are highly scattered. Davis *et al.* (1996b) show that multiple scattering leads to a non-trivial physical smoothing over scales  $\approx 200\text{--}300$  m, and Marshak *et al.* (1995b) use multifractal analysis to show that this smoothing affects large and small jumps in the horizontal distribution of LW path differently.

## 5. SEMI-EMPIRICAL CRITERIA (Statistical Interpretation of Scaling Regimes and Exponents)

In this section, we state or establish some results, typically inequalities between exponents, that enable us to classify data as stationary or not, ergodic or not, intermittent or not, according to the scaling behavior of various statistics. The same data can have conflicting attributes (e.g., stationary and nonstationary) as long as they refer to different ranges of scales. From a statistical standpoint, it will become apparent that stationarity, ergodicity, and intermittency are just different facets of the basic issue of data analysis: *What properties should we determine from our data? and How accurately can we estimate them from our finite sample?* From a physical standpoint, stationarity and intermittency are clearly more fundamental concepts than ergodicity.

### 5.1 Criterion for Stationarity (A Necessary Condition for Ergodicity)

In the Appendix, we describe procedures for synthesizing a number of scale-invariant models for stochastic processes and discuss their properties, among these “stationarity.” This body of theoretical knowledge is important for a variety of reasons. First, models enable validation of analysis software — “instrumental calibration” in our laboratory analogy — as well as sensitivity studies (how does the output depend on the amount and properties of the data being processed?). Second, models can be used in applications (e.g., cloud radiation studies). Last but not least, models generally have well-understood properties. Strictly speaking, stationarity is a property that can only be assigned to a model because the question is: *Are statistical quantities, as defined by ensemble averaging (i.e., over probability- or  $f$ -space), invariant under translation in  $x$ ?* This is not easy to address with data because we always operate with finite amounts of data; furthermore we rely heavily on spatial averages to estimate statistics in the first place, so at least some  $x$ -dependence is operationally erased.

Another theoretical question, more directly relevant to data analysis, is that of “ergodicity:” *Do spatial averages of increasing length for a single realization converge to ensemble averages (over all possible realizations)?* In data analysis we generally make implicit ergodicity assumptions before computing spatial averages, i.e., we expect them to converge to something meaningful. We also assume implicitly that this definite number we are seeking does not depend on when we start computing it (stationarity). It is therefore important to have guidelines as to what quantities are statistically well-defined, not just computable by some given algorithm.

Without exception, the scale-invariant models in the Appendix with

$$\beta < 1 \quad (28a)$$

are stationary in the "broad" sense where  $G(r) = \langle [f(x+r) - \langle f \rangle][f(x) - \langle f \rangle] \rangle$  depends only on  $r$ , we recall that, in this theoretical context, the 2nd-order autocorrelation function is obtained by averaging over all possible (or at least many)  $f$ 's, holding  $x$  constant. Davis *et al.* (1994b, 1996a) give more general arguments for drawing the line between stationary and nonstationary behavior at  $\beta = 1$  for scale-invariant processes.

If we only have

$$\beta < 3 \quad (28b)$$

then the model has (broad-sense) stationary increments:  $\langle [f(x+r)f(x)] \rangle = G(r) + \langle f \rangle^2$ , and even  $\langle f(x) \rangle$ , may or may not depend on  $x$  but the structure function of order  $q = 2$ , namely  $\langle [f(x+r) - f(x)]^2 \rangle$ , does not. We propose to use the criteria in Eqs. (28a,b) for real world data-streams as well as for theoretical models. For data,  $\beta > 1$  means that spatial estimates of  $\langle f(x+r)f(x) \rangle$  are likely to vary from one realization (or portion of data) to the next. For data,  $\beta < 3$  means that spatial estimates of  $\langle [f(x+r) - f(x)]^2 \rangle$  are likely to be robust (i.e., invariant under addition of new data).

In this scheme, the FIRE LWC data ( $\beta \approx 1.4$ ) in Figs. 2a-e is nonstationary with stationary increments for scales from  $\approx 20$  m to the integral scale  $\approx 20$ -40 km. For larger scales, we have no spectral information in Fig. 4c but the leveling-off of the structure functions in Fig. 6a confirms this estimate of the integral scale. This is symptomatic of stationarity (increments cease to grow). The FIRE Landsat data ( $\beta \approx 2.0$ ) in Fig. 3 is also nonstationary with stationary increments from the radiative smoothing scale  $\approx 200$  m to  $\approx 10$  km. Were it not for saturation at maximal gray level 255, spectral flattening (transition to stationarity) would occur closer to the integral scale for the cloud LW path fluctuations, itself likely to be around that of LWC.

In contrast, the  $\epsilon(\eta; x)$  fields in Figs. 2a'-e' ( $\beta \approx 0.7$ ) are stationary in spite of their intense spikiness. This may seem surprising since local means,  $\epsilon(r; x)$  with  $r \gg \eta$ , will fluctuate wildly, depending on the strength and number of spikes that fall in the interval  $[x, x+r]$ . The strong variability of these fields therefore contrasts with the conventional wisdom about stationarity, essentially that (in the usage of time-series analysis)

$$\text{'temporal statistics do depend little on when they are gathered.'} \quad (*)$$

The occurrence of spikes of course perturbs strongly the local statistics and therefore violates this operational definition of stationarity. We prefer to think of property (\*) as a consequence of ergodicity (which is more restrictive than stationarity): if, in general, running temporal averages converge reasonably fast to their ensemble counterparts as the sample size increases, then clearly we need not worry about where we start cumulating. The two unstated assumptions in effect here are:

- 1) 1-point statistics are Gaussian-type, i.e., that only relatively small deviations from mean or modal values are anticipated;
  - 2) 2-point correlations are of short range, i.e., the integral scale is relatively small.
- Samples of relatively small length therefore provide enough data to obtain accurate estimates of 1-point and 2-point statistics of all orders.

In our outlook, stationarity should have no bearing at all on the intensity of the fluctuations; it should however have a strong impact on their “rapidity” (since the smaller the spectral exponent the more variance in the small scales). In short, we propose to drop the above assumption #1. Even in strongly intermittent cascade processes  $\epsilon(x)$  with  $\langle \epsilon \rangle = 1$ , spikes (where  $\epsilon(x) \gg 1$ ) do not prevent the signal from

‘having a well-defined mode,’ (†)

(which is naturally  $\ll 1$ ). Property (†) is, in our opinion, a better description of a stationary time-series, whether Gaussian/ergodic or not.<sup>1</sup>

## 5.2 The Onset of Sampling Problems (Trivial Ergodicity Violation by Finite Datasets)

In the above considerations, only 2nd-order statistics are used, and we assume that the data is ergodic for that type of statistic (dominated by relatively frequent events). Going to ever higher-order moments, extreme events gain more and more weight. In estimators based on a finite amount of data, a single event will eventually dominate. This can be sensed in scaling analyses when the exponent functions,  $\zeta(q)$  or  $K(q)$ , become linear in  $q$ . This is a gradual transition but one can nevertheless define  $q_S$  such that

$$\zeta'(q) \text{ or } K'(q) \approx \text{constant for } q > q_S.$$

Schertzer and Lovejoy (1992) derive an expression for  $q_S$  in the frame of singular measures. Now some singular cascade models for  $\epsilon(r;x)$  (e.g., “ $p$ -models” discussed in the Appendix) are immune to sampling problems yet their  $K(q)$ ’s become asymptotically linear in  $q$ , simply because there is a well-defined maximal event present in every realization. In practice, one can easily test the hypothesis of obtaining roughly the same  $\max_x \{\epsilon(r;x)\}$  for every realization. So, in principle, there is no risk of misinterpreting the observation of linear trends in  $K(q)$ , at least for models.

Using the FIRE LWC data, Figs. 6b for  $\zeta(q)$  and 9b for  $K(q)$  are quasi-linear for  $q \gtrsim 3$ , which is our estimate for  $q_S$ . The  $\zeta(q)$  in Fig. 7b for the FIRE radiance data is quite linear for small  $q$ ’s (see below) but does not show an asymptote for large  $q$ .

## 5.3 Criterion for Intermittency (Non-Trivial Ergodicity Violation by Non-Gaussian Processes)

Suppose we want to model data (or the geophysical field it samples) with the simplest possible scale-invariant stochastic process. A question naturally arises: ‘Is a non-Gaussian multifractal model strictly necessary or is a simpler monofractal one good enough?’ Focusing on  $q$ th-order structure functions in Eq. (16), we find mono- (or “simple”) scaling —namely,  $\zeta(q) = \zeta(1)q$ — as soon as the trivial dimensional reasoning, that is

$$\langle |f(x+r) - f(x)|^q \rangle \approx \langle |f(x+r) - f(x)| \rangle^q, \quad (29a)$$

<sup>1</sup>A random sample is, by definition, dominated by modal values and will generally not give good estimates of high-order statistics, even of the mean if the distribution is sufficiently broad or skewed.

makes accurate predictions. Similarly, for singular measures, we would have

$$\langle \varepsilon(r;x)^q \rangle \approx \langle \varepsilon(r;x) \rangle^q, \quad (29b)$$

hence  $K(q) \approx K(1)q \equiv 0$ . We have argued above that, for all practical purposes (i.e., a finite amount of data is available), Gaussian-type increments  $f(x+r)-f(x)$  or  $\varepsilon(x)$  fields will indeed yield good estimates of relatively high-order moments that obey (29a,b). This type of data can be deemed “ergodic” and is necessarily non-intermittent in the sense that  $K(q)$  is vanishingly small. It is therefore important to give a quantitative meaning to the “ $\approx$ ” symbols in Eqs. (29a,b)?

In theory, fractal and multifractal properties require a small-scale limit ( $r \rightarrow 0$ ) to become mathematically precise. In practice, multifractal analyses are performed with finite amounts of data with a finite range of scales. There will almost<sup>1</sup> always be trivial ergodicity violations based on extreme events, as described briefly in the previous subsection. In turn, these events lead to a small but finite degree of multiscaling. *Can we distinguish between this spurious multifractality and its “real” counterpart (i.e., that is likely to be robust under addition of new data)?*

This exercise in ergodicity verification normally requires either obtaining more data or subdividing the available data and re-doing the analyses in order to monitor the effect of sample-size. Aurell *et al.* (1992), Marshak *et al.* (1994), Eneva (1994), and others have explored this sampling issue analytically or numerically with specific models that yield  $K(q) \equiv 0$  in the small-scale limit but  $K(q) \neq 0$  for a finite range of scales. Grivet-Talocia (1995) and others have investigated finite-size effects that cause stationary scaling processes with  $\beta < 1$  to have small but finite  $\zeta(1)$  and  $\zeta(2)$ , in contradiction with Eq. (18), let alone  $\zeta(q) \equiv 0$ . A general (model-independent) first-order answer to the above question is now derived from the hypotheses in Eqs. (27a,b).

Let  $\xi(r;x) = |f(x+r)-f(x)|$  or  $\varepsilon(r;x)$ , depending on the multifractal approach of interest. Equations (29a,b) become

$$\langle \xi(r;x)^q \rangle^{1/q} \approx \langle \xi(r;x)^{q_{\text{ref}}} \rangle^{1/q_{\text{ref}}}. \quad (30)$$

For specificity, we can take  $q_{\text{ref}} = 1$  but in some applications  $q_{\text{ref}} = 2$  would be a natural choice. In the following, it is assumed that some range  $[q_{\text{min}}, q_{\text{max}}]$  is explored for a number of scales, ranging from  $r_{\text{min}}$  to  $r_{\text{max}}$ . To make the “ $\approx$ ” sign in Eqs. (29–30) quantitatively meaningful, we just need to make sure that there is minimal information in the prefactors on  $q$ -dependence. There often is (cf. Figs. 6a and 7a), but of a trivial kind, just reflecting a poor choice of units. It is easy to select physical units for  $\xi$  that remove this trivial dependence for singular measures: by normalizing  $\varepsilon(r;x)$  so that the  $r$ -independent statistic  $\langle \varepsilon(r;x) \rangle$  is unity (cf. Fig. 9a). For structure functions, there is no  $r$ -independent case but we can always choose units that make  $\langle |f(x+r_{\text{max}})-f(x)|^{q_{\text{ref}}} \rangle = 1$ ; then, if there is little statistical information on  $q$  in the prefactors, the same should be approximately true in general:  $\langle |f(x+r_{\text{max}})-f(x)|^q \rangle \approx 1$ . Visual inspection of log-log plots of  $\langle \xi(r;x)^q \rangle$  vs.  $r$  is enough to show that this is possible: the intercepts of regression lines should be approximately linear in  $q$ .

<sup>1</sup>We exclude models that are cunningly ergodic in the sense of one or the other of the multiscaling statistics (cf. “ $p$ -model” in the Appendix for singularity analysis).

Using the definitions in Eqs. (16) or (20), we can rewrite the constraint for monoscaling in Eq. (30) as  $(r/r_{\max})^{A(q)} \approx (r/r_{\max})^{A(q_{\text{ref}})}$  where

$$A(q) = \begin{cases} \zeta(q)/q = H(q) \\ -K(q)/q = \tau(q)/q + (1/q-1)d = (1/q-1)[d-D(q)] \end{cases} \quad (31)$$

In the (lower) singular measures case,  $A(q)$  is a new non-increasing function.

Now we must decide at what degree of discrepancy in Eq. (30) we are compelled to take multiscaling into account. There is no universal answer for this at present; in particular, the number of samples will be a factor. However, there is a general consensus that multifractals are a framework for modeling “strong” variability. It seems reasonable to require that the ratio between both sides of Eq. (30) be less than, say, one order-of-magnitude or more at the opposite end of the range of scales from where their quantitative agreement can be imposed arbitrarily, simply by using the proper units. We did this at  $r = r_{\max}$ , so we will now focus on  $r = r_{\min}$ . Conversely, we can require that *bone fide* multiscaling data obey

$$\langle \xi(r; x)^q \rangle^{1/q} / \langle \xi(r; x)^{q_{\text{ref}}} \rangle^{1/q_{\text{ref}}} \Big|_{r=r_{\min}} \approx (r_{\min}/r_{\max})^{A(q)} / (r_{\min}/r_{\max})^{A(q_{\text{ref}})} \gtrsim B \quad (32)$$

where  $B$  is an arbitrary but relatively “big” number, like 10 or 100 (depending on the application, amount of data, etc.). Taking logs in Eq. (32) yields

$$\log \left( \frac{r_{\max}}{r_{\min}} \right) A(q) \Big|_q^{q_{\text{ref}}} \gtrsim \log B. \quad (33)$$

The chance of passing this test is clearly greater if (i) the range of scales  $r_{\max}/r_{\min}$  is increased, or (ii) the range of  $q$ 's is increased.

We now restrict our attention to singular measures with which we characterize intermittency routinely. The criterion in Eq. (33) can be translated into another for  $C_1$  if we make an assumption on  $A(q)$ , hence  $K(q)$ . For small enough values of  $q$ , the log-normal model,  $K(q)/C_1 = q^2 - q$  (cf. Appendix), is a reasonable fit to our generally parabola-shaped  $K(q)$  curve (e.g., Fig. 9b); we can therefore take  $A(q)/C_1 = 1 - q$ . Combining this simple expression with the inequality in (33) and maximizing the range of  $q$  (i.e.,  $q = q_{\max}$ ,  $q_{\text{ref}} = q_{\min}$ ), this  $A(q)$  function yields a simple—erring somewhat on the conservative side—criterion where we can isolate  $C_1$  as the only data-dependent quantity, the others being essentially instrumental: the data is “truly” multifractal only if

$$C_1 \gtrsim \frac{\log B / \log(r_{\max}/r_{\min})}{q_{\max} - q_{\min}}. \quad (34)$$

We have  $r_{\max}/r_{\min} \approx 10^3$  for most of the entries in Table 2. The denominator in the r.h.s. is 5; taking  $B = 10$ , we require  $C_1 \geq 0.07$  to qualify the data as truly multiscaling. One case (FIRE LWC on 7/14) fails the multifractality test; being so exceptionally smooth, this is not a surprise. The others cases pass, although one just border-line.

We noted already that the  $\zeta(q)$  results in Fig. 7b for the FIRE radiance fields are almost linear in  $q$  for small enough  $q$  (say, 0 to 2.5). Can we argue for multifractality with this data? In this case, we have  $r_{\max}/r_{\min} \approx 10^2$  (this is a somewhat conservative



estimate because of the large-scale effect of radiometry saturation). Values of  $A(q)$  for  $q_{\text{ref}} = 1$  and  $q_{\text{max}} = 5$  are  $A(1) = \zeta(1) = H_1 \approx 0.5$  and  $A(5) = \zeta(5)/5 \approx 2.0/5$ . Using base 10 logarithms, the l.h.s. of Eq. (33) is therefore  $\approx 2^+ \times (0.5 - 0.4) = 0.2^+$ , much less than the r.h.s. for  $B = 10$ . This argues for a monofractal model for this data. Surprising result since we have good evidence of multifractality in the associated LWC fields which are representative of the fluctuations of the extinction (photon scattering probability per unit of length) in the cloud. Numerical simulations by Marshak *et al.* (1995b) explain this paradox: multiple scattering processes smooth the large jumps in extinction more effectively than the small ones; the latter determine  $H_1$  and the former, the higher-order moments of the increments.

## 6. THEORETICAL CONSIDERATIONS (Geophysical Data Analysis as a Problem in Statistical Physics)

Experimentation is generally conducted in the laboratory with a theoretical model in mind, a hypothesis to test. At the level of generality that we have adopted in our survey of multifractal data analysis techniques, the most pressing theoretical questions are: ‘*Why is scaling almost universally observed in geophysical signals and fields?*’; and ‘*What can we learn about the underlying physical processes from the scaling properties?*’.

### 6.1 Thermodynamical Interpretation of Multifractal Quantities

In §3.3 we likened the statistical parameter  $q$  with a standard one in experimental work, namely ambient temperature. In the same vein, there is an increasingly popular interpretation of all multifractal exponents as thermodynamical quantities, first explored by Feigenbaum (1987) and recently surveyed by Muzy *et al.* (1994). In particular, diverging moments (important for modeling ergodicity/sampling problems) are perceived as 1st- or 2nd-order phases transitions, their signature being a discontinuity in the 1st- or 2nd-order derivatives of the “equilibrium curves”  $\zeta(q)$  or  $K(q)$ ; see Schertzer and Lovejoy (1992) for a discussion in the frame of singular measures.

There are solid physical reasons for exploiting this formal analogy. Indeed, the current rationale for using scaling analysis is that geophysical systems by nature have very many interacting degrees-of-freedom. We can think of the number of computational cells required to solve the coupled PDEs for Navier-Stokes equations, generally with ancillary constraints, for very high Reynolds numbers. In this respect, we imagine the generally large, complex geophysical system under consideration as a thermodynamical one: it has many allowable configurations in the sense that all the microscopic variables can each take on a number of values. Two classic examples are: positions, linear- and angular momenta of molecules in a gas; spin values of atoms or domains in a magnetic material. This defines a vast probability space, impossible to describe in any kind of detail. However, the macroscopically observable quantities are few (akin to temperature, free energy, entropy, magnetic induction, etc.) and are defined by ensemble-averages over all possible microscopic configurations. These “observables” generally depend little

on the detailed dynamics of the system and thus define universality classes (e.g., real gases, Ising models). The counterpart of the thermodynamical limit in statistical physics (i.e., very many interacting particles) here is the limit of a huge computational grid (hence a large range of scales) and we again expect some kind of universal behavior to arise.

## 6.2 Information Created by Breaking the Scaling Symmetry

Scaling can be viewed as a symmetry (or invariance) obeyed by the macroscopic geophysical system probed during data collection. In this case, we are dealing with an invariance under change of scale. There are other possible symmetries: “stationarity” (§5.1), invariance under translation; and “isotropy” (§3.1), invariance under rotation in 2D, time reversal or parity ( $x \rightarrow -x$ ) in 1D. Of course these symmetries are all of a statistical nature since exactly translationally and rotationally symmetric fields are constant. Generally speaking, the more symmetric the system, the less information is required to describe it. In our case, similar statistical properties for a whole range of scales are described by a single exponent (and a prefactor). It can be argued that information about any system can be gained only by breaking its symmetry. For instance, to measure the circumference of a circle, a mark must be placed somewhere (thus breaking its rotational symmetry). The first and last points in a time-series are special with respect to translations (a degree of nonstationarity is therefore introduced).

Our experimentation with cloud data has confirmed this. In discussing Table 2, we highlighted the similarity of  $q = 1$  multifractal exponents for the same type of cloud (marine Sc) at three different locales (FIRE data from the N-E Pacific, ASTEX data from the Mid-Atlantic, SOCEX data from the S-W Pacific). So a “bi-fractal” characterization tells us nothing about the local climatology; it does tell us however something about the universality of the thermo-hydrodynamical processes that shape the internal structure of marine Sc layers. In contrast, the scaling range, which is defined by scale-breaks at either end, does vary; it appears to be roughly in proportion with the thickness of the boundary layer (Davis *et al.* 1996a, Marshak *et al.* 1996).

From the cloud radiative perspective, the most interesting scale-break is the one at  $\approx 200$  m in Figs. 5 and 7a relating to the reflected radiance fields of marine Sc. Indeed, there is no counterpart of this statistically robust feature in the LWC data measured inside the same type of cloud. Davis *et al.* (1996b) survey the literature on this scale-break in the energy spectrum and they describe the radiative smoothing mechanism that produces it, via multiple scattering. Marshak *et al.* (1995b) use multifractal methods to investigate its implications for cloud remote sensing. In both studies, the phenomenology of the Landsat scale-break is based on a numerical Green’s function analysis of horizontal photon transport, uncovering its dependence on geometrical cloud thickness and photon mean free path (corrected for the forward scattering): the scale-break occurs at the harmonic mean of these two fundamental scales.<sup>1</sup> Information about inherent cloud properties can therefore be extracted from the observation of the scale-break. Moreover,

<sup>1</sup>Interestingly, this finding is traceable to the effect of the non-illuminated cloud boundary at finite range on the photons’ random walks that are modeled by *nonstationary* Brownian motion (cf. Appendix).

it is hoped that in the near future laser and low-light detector technologies will be combined to observe directly cloud radiative Green's functions. This would enable robust and inexpensive estimations of cloud thickness and density, both important quantities for balancing the Earth's radiative budget, hence forecasting climate change.

## 7. Summary

We have presented a conceptual model for geophysical data analysis based on laboratory work. The "sample" being probed is the data, generally collected in the field: either time-series, 1D transects, or 2D images. The laboratory "instruments" are computer codes that process this data and output 2- or more-point statistics, our focus being on scale-dependent statistical quantities that convey information about spatial correlations. The "readings" of these instruments are analyzed on  $\log[\textit{statistic}]$  vs.  $\log[\textit{scale}]$  plots, seeking straight lines that are the signature of scaling (power-law) regimes. A number of remarkable exponents (log-log slopes) are discussed and several criteria are presented: *When is the datastream stationary? ... When does it have stationary increments? ... Do we suffer from sampling ("ergodicity") problems? ... Is there enough intermittency to call for an inherently non-Gaussian "multifractal" model, as opposed to a simpler Gaussian one with "monofractal" statistics?* Finally, laboratory exercises are almost universally designed to validate or challenge the prevailing theory about the structure and dynamics of the sample. Experimental results can therefore prompt new theory. We show how this proves true in both cases for the theory of cloud-radiation interaction, an important pre-requisite in climate theory and the remote sensing of cloud properties. We therefore conclude that multifractal scale-by-scale analysis is a powerful —yet currently underexploited— tool in geophysical research, well-suited for connecting theory and measurements in a broad range of applications.

## Appendix. SIMULATION AND CALIBRATION (Scale-Invariant Models: Stationary or Not, Intermittent or Not, Ergodic or Not)

An important, quasi-universal application of statistical data analysis, multifractal or other, is to constrain models that attempt to reproduce the data. So stochastic simulation tools are often developed in parallel with data analysis methods. This activity has both analytical and computational aspects. On the one hand, we need to write code that implement specific algorithms for generating random functions of one or two variables. On the other hand, we need to know, preferably in closed-form, the dependence of the statistical quantities of interest on the parameters of the model. In our experience, applications are two-fold:

- Stochastic cloud models have proved invaluable for investigating cloud radiation properties (e.g., reflectance, transmittance, and absorption of solar- and laser beams).
- Validation of data analysis algorithms: before applying them to real data, it is crucial to see how they respond to artificial data.

In the “laboratory” analogy for data analysis used in this paper, the “instrumentation” is a collection of computer programs that process the data-stream into statistics (large input, small output). Others will help fit nonlinear (e.g., power-law) models to the statistics in a more or less supervised manner. Just like real experimental procedures, these instruments must be calibrated with standard input, in our case, samples of data with known and controllable statistical properties.

In this Appendix, we present a comprehensive collection of theoretical models to perform this task. Simultaneously, the somewhat abstract concepts of stationarity and intermittency are made more palatable. On the one hand, we need both mono- and multiscaling nonstationary random *functions* with stationary increments to calibrate structure-function analysis. In section A.1, we recall how the simplest nonstationary processes are obtained from white noises in both ergodic and non-ergodic situations; in section A.2, we describe monoscaling fractional Brownian motions. On the other hand, we need stationary mono- and multiscaling random *measures* to calibrate singularity analysis procedures. In section A.3 such models are presented. In section A.4, we return to nonstationary functions to add two multiscaling models, both developed first for cloud studies, that complete the collection. In section A.5 finally, we summarize and display graphically connections between the different types of model.

Here again, we are dealing with a number of computer programs that synthesize data (small input, large output). This output is denoted  $w(x_m)$ ,  $f(x_m)$  or  $\epsilon(x_m)$  on a 1D grid of constant  $\ell$  and size  $M = L/\ell$ :

$$x_m = m\ell, \quad m = 0, \dots, M-1. \quad (\text{A0a})$$

Generalization to 2D grids,

$$(x_{m_1}, y_{m_2}) = (m_1, m_2)\ell, \quad (m_1, m_2) = [0, M-1] \otimes [0, M-1], \quad (\text{A0b})$$

are straightforward in most cases, at least if statistically isotropic models are acceptable. In one (more involved) case, the 2D construction is described in full detail (§A.3.5). Where convenient, we denote spatial dimensionality by

$$d = 1, 2. \quad (\text{A1})$$

Unless explicitly stated otherwise, we use units of length where

$$\ell = 1. \quad (\text{A2})$$

For each model, we describe the generation algorithm in sufficient detail for direct coding; the main statistical properties are expressed as a function of their parameters, with either short derivations or references to the literature, and their stationarity, intermittency and/or ergodicity properties are discussed.

## A.1 Nonstationarity in 1D, Running Sums of Stationary Processes

### A.1.1 Gaussian White Noise and Brownian Motion

Brownian motion (Bm), a.k.a. the Wiener-Lévy process, is defined as the integral of white noise (w.n.), i.e., the running sum of a sequence of *uncorrelated* “steps” in the

forward or backward direction: the steps are denoted  $w(x_m)$ ,  $m = 1, \dots, M$ . These numbers are typically drawn from a Gaussian distribution but Bernoulli<sup>1</sup> and Laplace<sup>2</sup> distributions are also used. Throughout this Appendix, we denote by  $N(\mu, \sigma)$  Gaussian<sup>3</sup> random variables of mean  $\mu$  and variance  $\sigma^2$  (standard deviation  $\sigma$ ). The important requirement is finite variance to invoke the central limit theorem (CLT). For simplicity, we also require the vanishing mean to eliminate systematic drifts; so  $w(x_m) = N(0, \sigma)$ .

Bm is therefore

$$f(x_m) = f_0 + \sum_{j=1}^m w(x_j), \text{ no sum for } m = 0. \quad (\text{A3})$$

If  $w(x_m)$  is Gaussian—or, more generally (CLT), in the limit  $m \gg 1$ — $f(x_m)$  is Gaussian, and so are its increments  $\Delta f(r; x_m) = f(x_{m+r}) - f(x_m) = \sum_{j=m+1}^{m+r} w(x_j)$ , where  $m = 0, \dots, M-r$  and  $r = 0, \dots, M$ , because sums of Gaussians are Gaussian (with cumulated means and variances). It is easy to see that  $\langle |\Delta f(r; x_m)|^q \rangle = \langle |\Delta f(r; x_0)|^q \rangle \propto \langle [f(x_r) - f_0]^2 \rangle^{q/2} = [\sigma^2 r]^{q/2}$ , where  $\sigma^2 = \langle w(x_m)^2 \rangle$  and<sup>4</sup>  $q > -1$ ; thus  $\zeta(2) = 1$  and, in general,  $\zeta(q) = q/2$  from the definition of structure function exponents in Eq. (16). We now contrast the properties of w.n. and Bm:

- Gaussian w.n. is the prototypical *stationary process*: all its statistics are independent of the position  $x_m$ . Absolute 1-point moments are invariant,  $\langle |w(x_m)|^q \rangle = \langle |w(x_1)|^q \rangle$  for  $m = 1, \dots, M$  and  $q > -1$  as is the 2-point autocorrelation function

$$\langle w(x_{m+r})w(x_m) \rangle \equiv \sigma^2 \delta(r), \quad (\text{A4a})$$

form  $m = 1, \dots, M-r$ . W.n. is not only stationary but *ergodic*: spatial averages over a single but large enough realization converge to the above ensemble averages, largely because of the Gaussian nature of  $w(x_m)$ . In practice, very good estimates can be obtained from relatively small samples; see Fig. A1a for a case with  $M = 1024$ .

- Bm is the prototypical *nonstationary* process with stationary increments. Indeed, its 1-point statistics depend explicitly on  $m$ : we may have  $\langle f(x_m) \rangle = \langle f_0 \rangle$  ( $= 0$  in the following) for all  $m$  but<sup>5</sup>  $\langle f_m^2 \rangle = \langle f_0^2 \rangle + m \langle w_1^2 \rangle$  for  $m \geq 0$ . The same is true of the 2-point autocorrelation:

$$\langle f(x_{m+r})f(x_m) \rangle \propto \langle f_0^2 \rangle + |m+r| + m-r \quad (\text{A4b})$$

which follows from the identity  $2f(x_{m+r})f(x_m) = f(x_{m+r})^2 + f(x_m)^2 - [f(x_{m+r}) - f(x_m)]^2$ . In

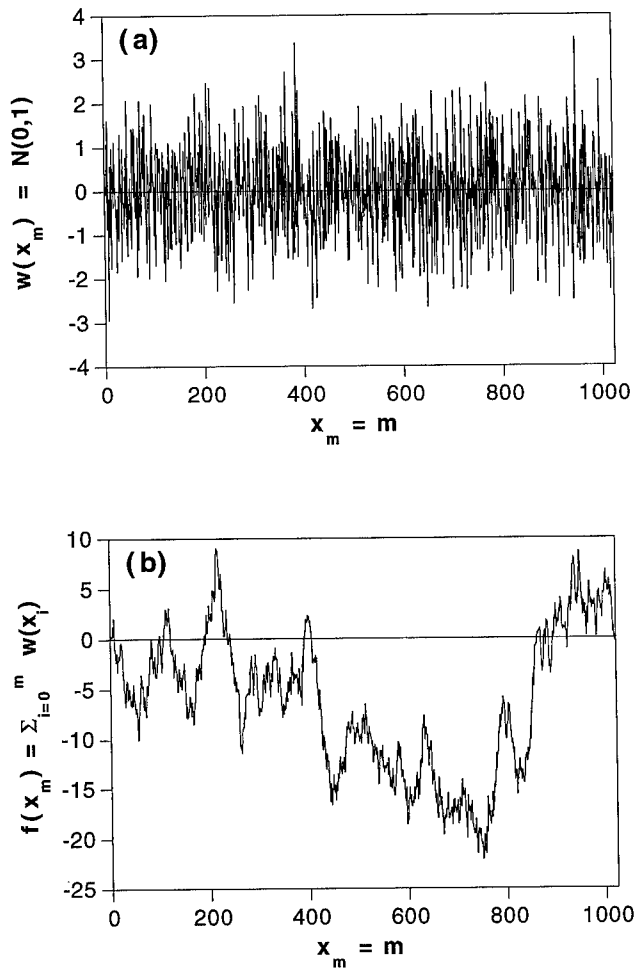
<sup>1</sup>A symmetric Bernoulli trial yields  $w(x_m) = \pm s$  with equally probable signs, hence Bm on a grid (widely used in diffusion theory).

<sup>2</sup>In neutron or photon transport, Laplacian or two-sided exponential free paths (mean  $\lambda$ ) are in order:  $\pm \lambda \ln(\xi)$  with  $\xi$  uniformly distributed on  $(0, 1)$  and equally probable signs.

<sup>3</sup>The Box-Muller transformation can be used to obtain zero-mean, unit-variance deviates; one way to do this is  $N(0, 1) = \cos(\pi \xi_1) (-2 \ln \xi_2)^{1/2}$  where  $\xi_i$  ( $i = 1, 2$ ) are computer-generated (pseudo-random) deviates distributed uniformly on  $(0, 1)$ . From there  $N(\mu, \sigma) = \mu + \sigma N(0, 1)$ .

<sup>4</sup>Generally the  $r$ -dependent pdf,  $\text{Prob}\{\Delta \leq \Delta f(r; x) < \Delta + d\Delta\}/d\Delta$ , of the increment  $\Delta f(r; x) = f(x+r) - f(x)$  ( $r \geq 0$ ) is non-vanishing at  $\Delta = 0$ ; moments  $\langle |\Delta f(r; x)|^q \rangle = \int \Delta^q \text{Prob}(d\Delta)$  are therefore divergent for  $q \leq -1$ .

<sup>5</sup>Setting  $\langle f_0^2 \rangle = 0$  enforces the nonstationarity (the  $m = 0$  point is special). With  $\langle f_0^2 \rangle = M \langle w_1^2 \rangle$ , detrending and cyclically extending it, the resulting Bm is (cyclo-)stationary. However, we are interested primarily in scales  $r \ll M$  for which this Bm is still nonstationary in the sense of the criterion in Eq. (28a):  $\beta = 2 > 1$ .



**Figure A1: Ergodicity, With and Without Stationarity.** (a) Uncorrelated (white) zero-mean unit-variance Gaussian noise; this data is not only stationary, but ergodic as well: spatial estimates of the low-order moments yield  $\mu = 1.4 \times 10^{-4}$ ,  $\sigma = 0.9962$  for  $M = 1024$ . (b) Running sum of the data in panel (a), i.e., Brownian motion; this data is nonstationary with stationary (and ergodic) increments.

contrast, 2-point statistics based on  $\Delta f(r; x_m)$  depend only on  $r$ . Turning to spatial averages on a scale of  $r$  pixels ( $1 \ll r \ll M$ ), 1-point means and variances will fluctuate wildly (non-ergodicity) but averages using  $\Delta f(r; x_m)$  are well-behaved (ergodic increments) as long as Gaussian-type distributions are used. In summary, Bm is neither stationary nor ergodic *per se* but has stationary and ergodic increments. For an illustration, Fig. A1b shows the running sum of the data in Fig. A1a.

### A.1.2 Lévy-Stable White Noise and Lévy Flights

White noise need not be Gaussian, hence so highly ergodic. Consider a power-law (e.g., Pareto) distribution where moments above some critical order are divergent; this means that their spatial estimates increase without bounds with sample size. Figure A2a illustrates this case with w.n. obeying a Cauchy law: moments of order  $q \geq 1$  diverge. Only absolute moments of very small non-integer order can be properly estimated with  $M \approx 10^3$ . The counterpart of Bm for the Cauchian w.n. is graphed in Fig. 2b. This running sum of the data in Fig. 2a is a “Lévy-sable” process, often referred to as a Lévy “flight” (Mandelbrot 1983). Here incremental ergodicity is of course violated.

By definition, Lévy-sable variables are a four-parameter class of deviates  $L(\alpha; a, b, c)$  obeying the rescaling equation:<sup>1</sup>

$$\sum_1^n L - na \stackrel{d}{=} n^{1/\alpha} (L - a), \quad (\text{A5})$$

- $\alpha$  is the “Lévy index.” For  $\alpha \rightarrow 0^+$ , the solution is  $L = a$  (the degenerate distribution) and normally distributed random variables obey the rescaling in Eq. (A5) when  $\alpha = 2$ . In general,  $0 < \alpha < 2$  is the critical order above which statistical moments diverge —variance included. For further details, we refer to Feller’s (1971) treatise.
- $a$  is a “centering” parameter, the counterpart of mean  $\mu$  in Gaussian deviates.
- $b \in [-1, +1]$  controls skewness and has no equivalent in the Gaussian case.
- $c$  is the amplitude parameter, like standard deviation  $\sigma$  in the Gaussian case.

We focus here on the most straightforward generalizations of the Gaussian case ( $b = 0$ ); denoting these  $L(\alpha; a, c)$ , we have  $N(\mu, \sigma) = \lim_{\alpha \rightarrow 2} L(\alpha; \mu, \sigma)$ . The only case with a closed-form pdf is  $\alpha = 1$  (Cauchy deviates); the sample in Figs. A2a,b also has  $a = 0$ .

### A.1.3 Spectral and Multifractal Properties

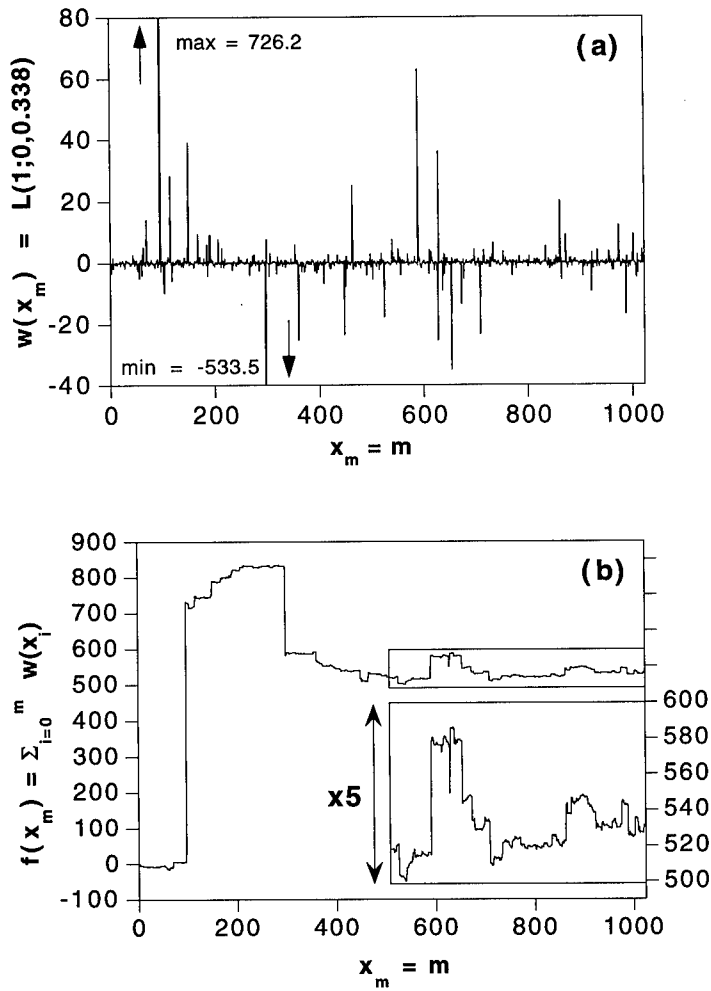
We have already shown that Bm (Gaussian-type increments) is monoscaling with

$$H(q) \equiv 1/2, \quad q > -1, \quad (\text{A6a})$$

The spectral exponent in  $E(k) \propto k^{-\beta}$  is related to the  $q = 2$  case by  $\beta = 2H(2) + 1 = 2$ .

The defining relation (A5) for non-skewed Lévy-stable deviates with vanishing mean ( $a = 0$ ) says that  $|\Delta f(r; x_m)|^q \stackrel{d}{=} |\sum L(\alpha; 0, c)|^q \stackrel{d}{=} r^{q/\alpha} |L(\alpha; 0, c)|^q$ . So, as long as  $-1 < q < \alpha$ , Lévy flight increments obey  $|\Delta f(r; x_m)|^q \propto r^{q/\alpha}$ ; therefore  $\zeta(q) = q/\alpha$ . What if  $q \geq \alpha$ ? In theory, the corresponding moments do not exist, so neither will the exponent  $\zeta(q)$ . In

<sup>1</sup>The symbol “ $\stackrel{d}{=}$ ” means “identity in distribution.”



**Figure A2: Non-Ergodic Stationary and Nonstationary Processes.** (a) White Lévy-stable noise with the same order 1/2 moment as the Gaussian case in Fig. A1a; this data is stationary but non-ergodic (at least for moments of order  $q \geq 1$ ). Symmetric Lévy-stable deviates with zero mean  $L(\alpha;0,c)$  with a Lévy “index”  $\alpha = 1$  and a scaling parameter  $c = 0.338\cdots = \Gamma(3/4)^2/\Gamma(\pi/2)$  were used, obeying a Cauchy law:  $d\text{Prob}\{X \leq L(1;0,c) < X+dX\} = cdX/[\pi(c^2+X^2)]$ . (b) Running sum of the data in panel (a), in other words, a “Cauchian” Lévy-flight; this data is nonstationary with increments that are stationary but not ergodic. The inset is a vertical zoom into the second half of the sample, showing more “jumps” of lesser magnitude.



practice, finite-sample estimates of  $\langle |\Delta f(r; x_m)|^q \rangle$  become dominated by the most intense event in the w.n.  $w(x_m) = L(\alpha; 0, c)$ . The corresponding value of  $x_m$  is straddled by exactly  $r$  of the  $M-r$  ( $> r$ ) increments included in the spatial averaging operation; this leads to  $\langle |\Delta f(r; x_m)|^q \rangle \approx \max_m \{ |w(x_m)|^q \} r/M \propto r$ , hence  $\zeta(q) = 1$  for  $q > \alpha$ . In summary, Lévy-stable processes are “operationally” multiscaling with  $\zeta(q) = \min\{q/\alpha, 1\}$ ; hence

$$H(q) = \min\{1/\alpha, 1/q\}, q > -1 \quad (0 < \alpha < 2). \quad (\text{A6b})$$

Here again we find  $\beta = 2$  (operationally) although for radically different reasons than in Bm. To obtain Bm, we integrate ( $\beta \rightarrow \beta+2$ ) a sample of w.n. ( $\beta = 0$ ) with spectral density  $E(k) \approx \text{constant} < \infty$ . In Lévy flights, spectral density (like variance) diverges:  $E(k) \rightarrow \infty$  with sample size, via the prefactor. In a finite sample,  $f(x)$  has a finite number of discontinuities, hence  $\beta = 2$  as for Heaviside steps.<sup>1</sup>

## A.2 Fractional Brownian Motions as Nonstationary Monofractal Functions

### A.2.1 General Properties of Fractional Brownian Motion

Processes known as “fractional Brownian motion” (fBm) are, like Lévy flights, generalizations of Bm but in a different direction. They were introduced by Mandelbrot and van Ness (1968) to reproduce the correlation properties of otherwise Bm-like “steps” observed in numerous physical signals, for instance, turbulent velocity. Like standard Bm (uncorrelated steps), fBm is a nonstationary random function with stationary increments.

The noteworthy property of all scale-invariant nonstationary signals and models, fBm in particular, that describes incremental correlations is

$$\langle \Delta f(r; x+r) \Delta f(r; x) \rangle = [2^{2H(2)-1} - 1] \langle \Delta f(r; x)^2 \rangle. \quad (\text{A7})$$

This follows directly from Eq. (16), expressed for  $r$  and  $2r$  at  $q = 2$ , and the stationarity of the increments. So, at  $H(2) = 1/2$  we retrieve Bm with characteristically uncorrelated increments in (A7). For  $0 \leq H(2) < 1/2$ , we have *negatively* correlated increments: a jump up is more often than not followed by a jump down and vice-versa. This leads to a *less* nonstationary process than Bm, “anti-persistence” in Mandelbrot’s (1983) words. For instance, Kolmogorov (1941) scaling in turbulent signals corresponds to  $H(2) \approx 1/3$ . For  $1/2 < H(2) \leq 1$ , we have the opposite situation: *positively* correlated increments or “persistence” in Eq. (A7). If a jump tends to be followed by another in the same direction, then we have *more* nonstationarity than in Bm. Mandelbrot notes that the Earth’s topography is reasonably well modeled by setting  $H(2) \approx 0.75$ .

The higher-order structure functions of fBm obey

$$H(q) = \zeta(q)/q \equiv H_2, \quad 0 < H_2 < 1. \quad (\text{A8})$$

We recall from §4.2 in the main body that relation (A8) arises whenever the increments are distributed narrowly enough, Gaussian-style, that we can use the simplest dimensional

<sup>1</sup> The only difference in Fourier space between w.n. and a Dirac  $\delta$  is the random phases; both have  $\beta = 0$ .

arguments to estimate higher-order moments from low-order ones: taking  $q = 2$  as the standard, then  $\langle |f(x+r) - f(x)|^q \rangle \approx \langle |f(x+r) - f(x)|^2 \rangle^{q/2}$ ,  $q > -1$ . The definition of structure function exponents in Eq. (16) then leads to  $\zeta(q) = \zeta(2)q/2 = qH(2) = qH_2$ , hence (A8).

Equation (A8) shows that fBm belongs to the restricted class of *monofractal* random functions; its only parameter relevant to scaling,  $H_2$ , is equal, in particular, to the mean Hölder exponent  $H_1 = H(1)$ . Figure A3a shows  $\zeta(q)$  versus  $q$  for  $H_1 = H_2 = 1/3, 1/2, 2/3$ , hence an increasing degree of nonstationarity. The two extreme cases are also shown:

- $H_1 = H_2 \rightarrow 0$  is the “stationary” limit where increments are not only scale-invariant but scale-independent; and
- $H_1 = H_2 \rightarrow 1$  is the “differentiable” limit where increments become (almost) everywhere directly proportional to the scale.

In the remainder of this section, we describe the two major computational routes to fBm—one in physical space, the other in Fourier space—and a variant using running sums. The common denominator of these methods is to use only Gaussian random numbers and additions, hence the term “additive” is frequently used for this whole class of models. Sums of normal deviates are normally distributed, so all quantities involved are Gaussian—most importantly,  $f(x)$  and  $f(x+r) - f(x)$ —therefore Eq. (A8) is verified by construction. Non-Gaussian multifractal models that violate (A8) were discussed above and will be again in §A.4, after surveying the prerequisite singular measures in §A.3.

#### A.2.2 Synthesis in Physical Space (Recursive Mid-Point Displacement)

The simplest algorithm for making fBm is known as “mid-point displacement,” cf. Peitgen *et al.* (1988). For procedures in physical space, it is generally convenient to use units of length where the grid constant is unity and we take a power of 2 plus one grid size:

$$M = L_n + 1 = 2^n + 1, \quad n > 0. \quad (\text{A9})$$

To determine  $f_n(x_j)$  for  $j = 0, \dots, 2^n$ , we first set<sup>1</sup>

$$f_n(0) = f_0 \quad (\text{A10})$$

$$f_n(L_n) = \begin{cases} N(f_0, \sigma_0) \text{ where } \sigma_0 = 2^{(1-H_2)} - 1, \text{ or} \\ f_n(0) \text{ for cyclical boundary conditions} \end{cases} \quad (\text{A11})$$

This completes the maximum scale  $r_0 = L_n$ . We then proceed recursively to smaller and smaller scales,

$$r_i = r_{i-1}/2 = r_0/2^i = 2^{n-i} \quad (i = 1, \dots, n), \quad (\text{A12})$$

using a decreasing sequence of standard deviations,

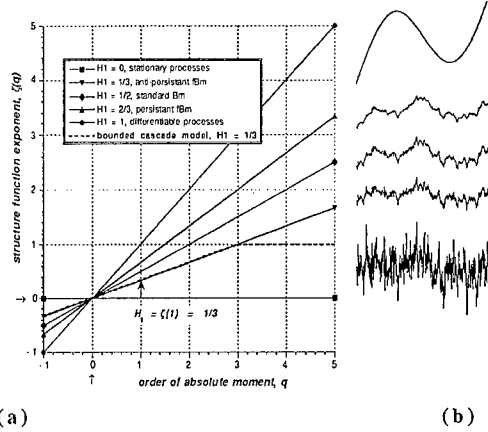
$$\sigma_i = \sigma_{i-1}/2^{H_2} = \sigma_0/2^{iH_2} \quad (i = 1, \dots, n). \quad (\text{A13})$$

New points are generated by averaging existing ones and adding random displacements:

$$f_n(x_j + r_i) = [f_n(x_j + 2r_i) + f_n(x_j)]/2 + N(0, \sigma_i), \quad x_j = 2^j r_i \quad (j = 0, \dots, 2^i - 1). \quad (\text{A14})$$

<sup>1</sup>The value of  $f_n(0)$  in (A10) is arbitrary but deterministic, hence the patently nonstationary nature of fBm.

Results for five cases are presented in Fig. A3b, illustrating more and more nonstationarity:  $H_2 = 0$  ("1/f" noise, cf. next section);  $H_2 = 1/3$  (anti-persistence, as in turbulent signals);  $H_2 = 1/2$  (Bm);  $H_2 = 2/3$  (persistence, as in topography); and  $H_2 = 1$  (a noiseless linear trend). For the generalization to 2D, we refer to Peitgen *et al.* (1988).



**Figure A3:** Fractional Brownian Motions in 1D, with More or Less Nonstationarity. (a)  $q$ th-order structure function exponents  $\zeta(q) = qH_1$ . (b) Three different fBm's ( $H_1 = H_2 = 1/3, 1/2, 2/3$ , by increasing degree of nonstationarity) and two limiting cases:  $H_1 = 0$  ("1/fnoise," border-line stationarity) and  $H_1 = 1$  (differentiability, maximum nonstationarity on this scale).

### A.2.3 Synthesis in Fourier Space (Power-Law Filtering)

Another way of generating fBm is with low-pass power-law filtering in Fourier space, a procedure also known as "fractional integration" that we will invoke again further on to generate multifractal signals.<sup>1</sup>

Following Voss (1983), we start with Gaussian white noise on a grid of size  $2^n$ :

$$w_n(x_j) = N(0,1), j = 0, \dots, 2^n - 1. \quad (\text{A15})$$

This trivial stochastic process made of normal deviates, completely uncorrelated from one grid point to the next, is scale-invariant by construction and characterized by a spectral exponent

$$\beta_w = 0. \quad (\text{A16})$$

We now want to incorporate correlations, leading to

$$\beta = 2H_2 + 1. \quad (\text{A17})$$

This is easily done by power-law filtering in Fourier space, where it is convenient to set

$$L = 2^n \ell_n = 1. \quad (\text{A18})$$

<sup>1</sup>Pearson (1990) surveys fractional integration, as defined in the mathematical literature, which is somewhat different from ours.

After obtaining<sup>1</sup>

$$\begin{cases} \tilde{f}_n(0) = \tilde{w}_n(0) \approx 0 \\ \tilde{f}_n(k) = \tilde{w}_n(k) \times |k|^{-(H_2+1/2)}, k = \pm 1, \dots, \pm 2^{n-1}, \end{cases} \quad (\text{A19})$$

we compute  $f_n(x_j)$  by inverse Fourier transform for  $x_j = j\ell_n$  ( $j = 0, \dots, 2^n - 1$ ). The exponent  $H_2 + 1/2$  ( $0 < H_2 < 1$ ) in Eq. (A19) is naturally the “order of the fractional integration.” Generalization to 2D is straightforward.

Equation (A17) still applies if we take  $H_2 < 0$  hence  $\beta < 1$ , or  $H_2 > 1$  hence  $\beta > 3$ , in (A19). The latter choice yields nonstationary random functions with *nonstationary* increments, varying so weakly in the small scales that they are everywhere continuous and almost everywhere differentiable. So, increments become proportional to distance between points of interest:  $\langle |f(x+r) - f(x)| \rangle \approx \langle [f(x+r) - f(x)]^q \rangle^{1/q} \propto r$ , leading to  $H(q) \equiv 1$ . We now discuss the consequences of the former choice.

#### A.2.4 Synthesis from Stationary Gaussian Scaling Noises

What if we take  $H_2 < 0$  in the recipe contained in Eq. (A16)? The recursive mid-point displacement procedure in §A.2.3 can also be applied for negative<sup>2</sup>  $H_2$ . However, the outcome for  $H_2 < 0$  and  $H_2 > 0$  have radically different properties.

- If  $H_2 < 0$ , the corresponding scale-invariant noise has an energy spectrum  $E(k) \propto k^{-\beta}$  with  $\beta < 1$ ; there is finite energy (i.e., variance) at large scales ( $k \rightarrow 0$ ) but divergence at small ones ( $k \rightarrow \infty$ ) — a so-called “ultra-violet” catastrophe occurs. This means that small-scale singularities will develop whereas large-scale properties will remain relatively well-behaved.
- For  $H_2 > 0$  ( $\beta > 1$ ), the opposite occurs — an “infra-red” catastrophe. We can therefore expect large fluctuations at scales comparable to the computational domain, but relatively small ones at the pixel scale.

Only for  $H_2 < 0$  is there a well-defined, monotonically decreasing power-law autocorrelation function:<sup>3</sup>  $\langle f(x+r)f(x) \rangle \propto r^{-(1-\beta)} = r^{-2H_2}$ , hence stationarity in the broad-sense. For  $0 < H_2 < 1$ , a power-law structure function for  $q = 2$  follows from a power-law energy spectrum  $E(k)$ :  $\langle [f(x+r) - f(x)]^2 \rangle \propto r^{2H_2}$ , hence (broad-sense) stationary increments.

We will refer to these models with  $H_2 < 0$  as “Stationary Gaussian Scaling Noises” (or SGSNs). One-dimensional fBm with spectral exponent  $1 < \beta < 3$  can be obtained by standard integration (i.e., running sums in physical space) of SGSNs with  $\beta' = \beta - 2$  such that  $|\beta'| < 1$ . These SGSNs are first obtained by fractional integration of Gaussian w.n. at order  $\beta'/2 = H_2 + 1/2$ , ranging from  $-1/2$  to  $1/2$  ( $-1 < H_2 < 0$ ). We note that, for the anti-persistent case ( $1 < \beta < 2$ ), the corresponding SGSNs results in effect from a fractional

<sup>1</sup>Note that we leave the phases  $\tan^{-1}(\text{Im}[\tilde{w}_n(k)]/\text{Re}[\tilde{w}_n(k)])$  of the Fourier components unchanged. However, to retrieve Bm exactly (as defined in §A.1.1), we need to change the phases by exactly  $\pi/2$ :  $\tilde{f}_n(k) = \tilde{w}_n(k)/(ik)$  will yield a realization with  $f_0 = \tilde{w}_n(0)L \approx 0$ .

<sup>2</sup>The choice  $H_2 = -1/2$  leads to a rather convoluted way of generating uncorrelated values at every pixel. In contrast, this choice implies a null operation in the Fourier-based recipe in Eq. (A19).

<sup>3</sup>This requires  $0 < \beta < 1$  ( $-1/2 < H_2 < 0$ ), otherwise ( $H_2 < -1/2$ ) there are anti-correlations from one grid-point to the next; thus  $\langle f(x+r)f(x) \rangle < 0$ , and a power-law parameterization is invalid.

“differentiation” of Gaussian white noise, the order of “integration”  $\beta'/2$  being negative, between 0 (for Bm) and  $-1/2$  (for “ $1/f$ ” noise).

### A.3 Cascades Leading to Singular, Stationary, Mono- and Multifractal Measures

Calibration of singularity analysis algorithms calls for a different class of models than described above. The main requirement is that the model can be read as a “measure,” a non-negative<sup>1</sup> field that we will denote generically by  $\varepsilon(x)$ . This excludes the models discussed up to now. In practical data analysis applications, the nonstationary data-stream can be used to generate a measure, typically by taking small- but finite-scale gradients then their absolute values or squares. We can do this for the models presented in section A.2 but the results are quite disappointing, the small-scale gradients fields of Gaussian models being at best weakly variable (narrowly distributed).

Let  $f(x)$  be a nonstationary *additive* process based on  $w(x)$ , a Gaussian-type noise such as an SGSN. This gives  $\varepsilon(x) = |w(x)|$  which varies less than  $w(x)$  itself. Being essentially a running mean, the coarse-grained measure  $\varepsilon(r;x) = (1/r) \sum_{x'=x}^{x+r-1} |w(x')|$  will be statistically independent not only of  $x$ , stationarity oblige, but also of  $r$  over a large range of values because of the strong ergodic property of Gaussian-like  $w(x)$ . This translates to  $K(q) \equiv 0$  in Eq. (20), apart from finite sampling effects (Aurell *et al.* 1992). The same is true for models of  $w(x)$  with power-law tails, except that moments exist only up to some critical order  $q_D$ :  $K(q) \equiv 0$  for  $q < q_D$ ; for instance, Lévy-stable variables yield  $q_D = \alpha < 2$ .

To obtain models with  $K(q) \neq 0$ , we must leave the realm of *additive* models and generate measures with *multiplicative* cascades. In the same way as fBm's are monofractal random functions  $f(x)$  with multifractal counterparts (some of which are described in Section A.4), there are mono- and multifractal measures. The literature on multifractal measures is vast and increasing at a rapid pace. In this survey, we will start with some well-known monofractal cases in 1D and 2D (§A.3.1); straightforward generalizations of these (§A.3.2) lead to multiscaling in 1D (§A.3.3–4) and 2D (§A.3.5).

Mono- and multifractal cascade models come in two distinct flavors: the “canonical” (§A.3.3) and the “microcanonical” (§A.3.4–5). The latter, having the same singularity properties for every realization (ergodicity in the sense of singular measures), are particularly useful for the purposes of calibration; the former are arguably more realistic models for geophysical fields (Schertzer and Lovejoy 1987, Gupta and Waymire 1993). However, they are always non-ergodic at some level of confidence and can sometimes exhibit the interesting statistical phenomenon of divergence of higher-order moments. If ignored, this last characteristic can affect a calibration procedure.

#### A.3.1 Cantor's, Dirac's and Other Monofractal Measures

Consider the example of Cantor's measure, supported entirely by Cantor's famous set (Fig. A4a) which has fractal dimension  $D_f = \log 2 / \log 3 = \log_3 2 = 0.631\ldots < d = 1$ .

<sup>1</sup>Mathematically, “measures” can be thought of as “generalized” functions in the sense of Dirac (i.e., defined only under integrals) that are furthermore non-negative. The sum of integrals over disjoint sets is therefore equal to the integral over their union.

Cantor's measure is generated by starting with a uniform distribution,  $\epsilon_0(x) = 1$  on  $[0,1]$  which is then divided into 3 equal parts. The middle<sup>1</sup> third is emptied of its measure—this cell (or “eddy”) is now “dead”—and its mass is uniformly redistributed between its neighbors where  $\epsilon_0(x)$  therefore becomes  $\epsilon_1(x) = \epsilon_0(x) \times (1+1/2) = 3/2$ . We have thus required the spatial average to remain constant. After  $n$  steps, the scale is  $r_n = 3^{-n}$ , there are  $3^n$  cells in all, only  $2^n$  of which carry a measure  $\epsilon_n(x) = (3/2)^n$ , elsewhere  $\epsilon_n(x) = 0$ . Although deterministic, one can still define the spatial statistics of this measure:  $\langle \epsilon(r_n; x)^q \rangle = (2/3)^n \times (3/2)^{nq} = [(1/3)^n]^{(1-q)\log_{1/3}(2/3)}$ . Equation (20) in the main text then yields

$$K(q) = [1 - \log_3 2](q-1), q > 0. \quad (\text{A20a})$$

For  $q = 0$  we have  $K(q) = 0$  (by definition) and, for  $q < 0$ , we find  $\langle \epsilon(r_n; x)^q \rangle = \infty$  (formally, we can set  $K(q) = \infty$  here) because the empty events,  $\epsilon(r_n; x) = 0$ , dominate. We note that the factor  $1 - \log_3 2$  in (20a) is the codimension  $d - D_f$  of Cantor's set.

An example embedded in 2D is easily generated as follows. Start with a uniform distribution,  $\epsilon_0(x, y) = 1$ , on the unit square; then divided it into 4 equal parts; one sub-square is picked at random<sup>2</sup>, and its measure (1/4 of total) is set to 0; the measure in the neighboring cells is boosted proportionately, becoming  $\epsilon_1(x, y) = \epsilon_0(x, y) \times (1+1/3) = 4/3$ . After  $n$  steps, the scale of interest is  $r_n = 2^{-n}$ ; there are  $4^n$  cells in all with only  $3^n$  of them containing a non-vanishing measure,  $\epsilon_n(x, y) = (4/3)^n$ . Spatial averaging leads therefore to  $\langle \epsilon(r_n; x)^q \rangle = (3/4)^n \times (4/3)^{nq} = [(1/2)^n]^{(1-q)\log_{1/2}(3/4)}$ , hence

$$K(q) = [2 - \log_2 3](q-1), q > 0. \quad (\text{A20b})$$

The support of this measure is a fractal set of codimension  $-K(0^+) = 2 - \log_2 3$ , hence a fractal dimension  $D_f = 2 + K(0^+) = \log_2 3 = 1.585 \dots < d = 2$ , as can be verified by direct box-counting methods.<sup>3</sup>

The simplest random monofractal measure is probably a Dirac  $\delta$ -function positioned at a random point  $x^*$  on the unit segment in  $d = 1$ :  $\epsilon(x) = \delta(x - x^*)$ ,  $0 < x^* < 1$ ; we can also consider the unit square in the  $d = 2$  case:  $\epsilon(x, y) = \delta(x - x^*)\delta(y - y^*)$ ,  $(x^*, y^*) \in [0, 1]^2$ . These measures are entirely concentrated onto a single point; in other words the fractal dimension of their support is  $D_f = 0$ , hence a codimension  $d - D_f = d$ . Let us estimate the statistical moments of the coarse-grained measure  $\epsilon(r; x)$  contained in the sub-interval  $[x, x+r)$  or sub-domain  $[x, x+r) \otimes [y, y+r)$ ,  $0 \leq r \leq 1$ , according to Eq. (14): taking  $x$  (and, if necessary,  $y$ ) at random, we find  $\epsilon(r; x) = 1/r^d$  with probability  $r^d$  and 0 otherwise. Therefore, we have  $\langle \epsilon(r; x)^q \rangle = r^d / r^{dq} = r^{(1-q)d}$ ; hence, from Eq. (20):

$$K(q) = (q-1)d, q > 0. \quad (\text{A20c})$$

All of the above formulas (A20a,b,c) can be parameterized as

$$K(q) = (q-1)C_0, q > 0, \quad (\text{A21})$$

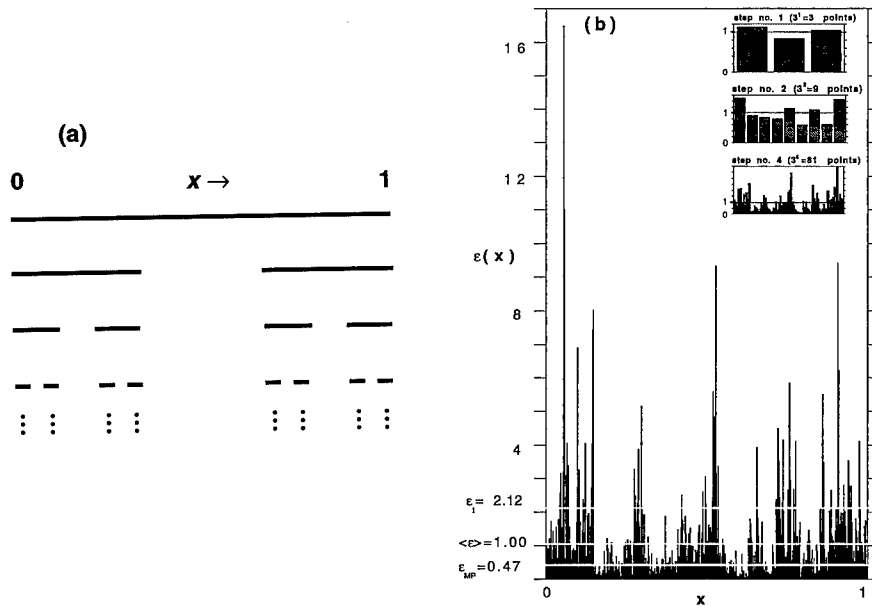
<sup>1</sup>The “middle-third” convention leads to a deterministic measure. The spatial statistics are unchanged by picking the next dead cell at random.

<sup>2</sup>If the same (e.g., upper-left) sub-square is picked each time, the limiting set is akin to Sierpinski's triangle.

<sup>3</sup>Total number of boxes of size  $r = 1/2^n$  ( $n \geq 0$ ) in unit square:  $N_t = (2^n)^2 = (1/r)^d$  ( $d = 2$ ). Number needed to cover fractal set:  $N_s = 3^n \propto (1/r)^{D_f}$ . Fractal dimension of set:  $D_f = \log N_s / \log(1/r) = \log_2 3$  (QED).

where  $C_0 = d - D_f > 0$  is the codimension of the support. The above measures are “monofractal” in the sense that their supports are nontrivial fractals; however, on these sparse subsets of space, they are uniformly distributed. Why their exponents  $K(q)$  are always in the linear relationship spelled out in Eq. (21) can be traced to this uniformity.

Another common feature of the above monofractal models is that they were constrained, largely for tutorial purposes, to have the same statistics for every realization (and indeed at every cascade step); in particular, the total measure is conserved. Such models are called “microcanonical” (Mandelbrot, 1974). “Canonical” counterparts of these models—where only the probability of killing a cell is prescribed, irrespective of what happens to its neighbors—have been proposed to model 1D transects of the kinetic energy dissipation field in fully-developed 3D turbulence (Novikov and Stewart, 1964; Mandelbrot, 1974; Frisch *et al.*, 1978). Indeed there is no reason to require a 1D sample of a 3D field to obey a conservation law that applies at best to the whole volume of the system. The observed statistics of turbulent flows are best reproduced by models with  $C_1$  values in the range 0.2–0.3. Following the nomenclature of Frisch *et al.* (1978), we refer to these monoscaling canonical cascade models collectively as “beta” models.



**Figure A4:** Two Multiplicative Cascades in 1D with branching ratio  $\lambda = 3$ . (a) First 3 cascade steps in the construction of Cantor's deterministic measure, with  $D(q) \equiv \log_3 2 = 0.63\cdots$ . (b) A random log-normal measure, cascade steps # 1, 2, 4 (inset) and 6, with  $D(q) = d - C_1 q$  where  $d = 1$ ,  $C_1 = 0.11\cdots$  and  $q < d/C_1$ .

### A.3.2 Canonical Random Measures and Divergence of Higher-Order Moments

There is no reason to limit ourselves to uniform measures. A simple way of obtaining non-uniform measures is to allow the random multiplicative “weights” used implicitly in the above cascade algorithms to differ from 0 or a constant. In general, we can envision a “turbulent” cascade process in  $d = 1$  or 2 dimensions that proceeds by divisions into  $\lambda^d$  “sub-eddies” of equal size at each step; the integer  $\lambda$  is known as the “branching ratio.” After  $n$  steps the scale is  $r_n = \lambda^{-n}$ , we have

$$\epsilon_n = \prod_{i=1}^n W_i, \quad (\text{A22a})$$

requiring only that

$$\langle W \rangle = 1. \quad (\text{A22b})$$

Redefining temporarily  $K(q)$  as  $-\ln \langle \epsilon_n^q \rangle / \ln r_n$ —as opposed to the parameterization in the main text’s Eq. (20) of the integrals (measures) defined in Eqs. (14–15)—it is easy to see that

$$K(q) = \frac{\ln \langle W^q \rangle}{\ln \lambda} = \log_\lambda \langle W^q \rangle, \{q \in \mathfrak{R}; \langle W^q \rangle < \infty\}. \quad (\text{A23a})$$

Defined in this way,  $K(q)$  inherits all of the analytical properties of cumulant-generating function<sup>1</sup> for the random variables  $\ln W$ . In particular,  $K(q)$  is convex with  $K(0) = 0$ . We also know that  $K(1) = 0$  due to (A22b).

Returning to the original definition of  $K(q)$  in relation to the measures in Eq. (20) is quite involved mathematically (Mandelbrot, 1974; Kahane and Peryière, 1976; Schertzer and Lovejoy, 1987; Gupta and Waymire, 1993). However, the only difference is ultimately that the trivial condition on  $q$  in Eq. (A23a) is replaced by

$$\{q < q_D; \frac{K(q_D)}{q_D - 1} = d\}. \quad (\text{A23b})$$

For  $q \geq q_D$ , the moments  $\langle \epsilon(r; x)^q \rangle$  are divergent. In practical data analysis applications, the symptom of a diverging moment is that its estimate is unstable, being dominated by the single largest event. The intensity of this event will depend critically on the sample size (in this case, grid size and number of realizations); we refer to Schertzer and Lovejoy (1992) for further details.

In summary, we can view the monofractal models of §A.3.1 as a limiting case in a continuum of models. The codimension of the measure’s support is<sup>2</sup>

$$C_0 = -K(0^+). \quad (\text{A24})$$

If, as in monofractal cases,  $C_0 > 0$  then it is arguably the most important parameter of the model since it defines geometrically the concentration of the measure on a sparse subset

<sup>1</sup>The characteristic function of a random variable  $\xi$  is defined as  $\phi(t) = \langle \exp[i t \xi] \rangle$ , i.e., the Fourier transform of its pdf; the cumulant-generating function (or “2nd characteristic function”) is  $\ln \phi(t)$ . Taking  $\xi = \ln W$  and  $it = q$ , we have  $K(q) = \ln \phi(q/i) / \ln \lambda$ . For imaginary arguments,  $\phi(t)$  is real and convex (Feller 1971).

<sup>2</sup>For  $q = 0^+$  everywhere  $\epsilon(r; x) > 0$  we have  $\epsilon(r; x)^q = 1$ , otherwise (i.e.,  $\epsilon(r; x) = 0$ ), we have  $\epsilon(r; x)^q = 0$ .



of space. In many cases however (cf. next sub-section), the probability of drawing exactly null weights is vanishingly small; this leads to  $K(0^+) = K(0) = 0$ , hence  $C_0 = 0$ .

In situations where  $C_0 = 0$  (i.e., the measure is supported by all of space), the next simplest way of quantifying its degree of concentration (hence intermittency) is the information codimension:

$$C_1 = K'(1). \quad (\text{A25})$$

This quantity can be obtained directly from the weights since Eqs. (A23a) and (A25) yield  $K'(1) = \langle W \log_\lambda W \rangle$ . For monofractal models, the linear formula for  $K(q)$  in Eq. (A21) yields  $K'(1) = K(0^+)$ , hence  $C_0 = C_1$ . We also note that,  $W \log_\lambda W$  being a convex function of  $W$ , Jensen's (1906) inequality tells us that  $C_1 = \langle W \log_\lambda W \rangle \geq \langle W \rangle \log_\lambda \langle W \rangle = 0$ .

Yet another way of parameterizing intermittency uses the correlation codimension

$$C_2 = K(2). \quad (\text{A26})$$

This 2nd-order statistic is the preferred choice in the turbulence literature, where  $C_2$  is in fact referred to as the "intermittency parameter" (and denoted " $\mu$ "). It determines the measure's spectral exponent<sup>1</sup>

$$\beta_\varepsilon = 1 - K(2) < 1. \quad (\text{A27})$$

The inequality follows from  $K(q)$ 's convexity and  $K(1) = 0$  which imply  $K(2) > 0$ . This establishes the stationarity of cascade processes according to the criterion in Eq. (28a).

### A.3.3 The Canonical Log-Normal Model, Stationarity in Presence of Intermittency

Consider a specific example: we use the Cantor measure construction ( $\lambda = 3$ ) with log-normal rather than Bernoulli weights. Instead of  $W_i = 0$  (Prob 1/3) or 3/2 (Prob 2/3), we take:

$$W_i = \exp[N(\mu, \sigma)] = \exp[\mu + \sigma N(0, 1)] \quad (i = 1, \dots, n), \quad (\text{A28a})$$

with 
$$\mu = -\frac{\sigma^2}{2}, \quad (\text{A28b})$$

to properly normalize the cascade,  $\langle W \rangle = 1$ .

In Fig. A4b, we illustrate the first steps of the construction using  $\sigma = 0.4$ . Notice the development of multiple singularities. At the same time, most of  $\varepsilon_n(x)$ 's values become smaller as  $n$  is incremented. Indeed, at constant  $x$ ,  $\ln \varepsilon_n(x) = \sum_0^n \ln W_i = \sum_0^n [\mu + \sigma N(0, 1)]$  is executing a random walk as a function of (discrete) "time"  $n$ , like the Bm described in §A.1.1 but with a systematic drift in the negative direction, due to (A28b).

Log-normal cascades were first introduced in turbulence theory by Kolmogorov (1962) and Obukhov (1962) to account for the effects of intermittency in the kinetic energy dissipation rate at high Reynolds numbers;  $\sigma$  around 0.7 was found to fit the data. We recall that, like SGSNs (§A.2.4), the model in Fig. A4b exemplifies stationarity according to criterion (28a) in the main text, cf. Eq. (A27). However, in sharp contrast with SGSNs, these cascades are highly non-Gaussian, not unlike the Lévy-stable w.n.

<sup>1</sup>This follows from  $\langle \varepsilon(x) \varepsilon(x+r) \rangle \sim r^{-K(2)}$  (Monin and Yaglom 1975, pp. 618–620) which is in Fourier duality with the energy spectrum (Wiener–Khinchin theorem)  $E_\varepsilon(k) \sim k^{-\beta_\varepsilon}$  with  $\beta_\varepsilon + K(2) = 1$ .

described previously. This is traceable to the multiplicative nature of the construction. In contrast with the white Lévy-stable noise, cascades have more interesting correlation properties due to the recursive nature of their construction.

Equations (A23a) and (A28a,b) yield<sup>1</sup>

$$K(q) = \frac{\sigma^2}{2\ln\lambda} q(q-1), \quad (\text{A29})$$

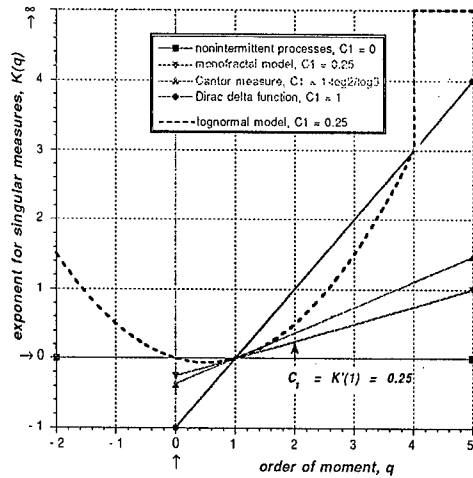
Using (A25), this can be rewritten<sup>2</sup>

$$K(q) = C_1 q(q-1), \quad q < q_D = d/C_1, \quad (\text{A30})$$

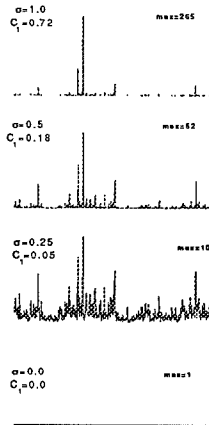
by identifying  $C_1$  with  $\sigma^2/(2\ln\lambda)$ . The condition on  $q$  in Eq. (A30) accounts for the divergence of moments, as predicted by Eq. (A23b).

In Fig. A5a, we have plotted  $K(q)$  for a log-normal model in  $d = 1$  with  $C_1 = 0.25$ ; notice the divergence of moments, formally  $K(q) = \infty$ , for  $q > q_D = 4$ . For comparison, we have also plotted  $K(q)$  for two of the monofractal models discussed previously, Dirac  $\delta$ 's and Cantor measures. Figure A5b shows a sequence of log-normal cascade models with increasing degrees of intermittency:  $\sigma = 0$  ( $C_1 = 0$ );  $\sigma = 0.25$  ( $C_1 \approx 0.05$ );  $\sigma = 0.5$  ( $C_1 \approx 0.18$ ); and  $\sigma = 1.0$  ( $C_1 \approx 0.72$ ). Notice the increasing concentration and spikiness.

**Figure A5: Log-Normal Cascade Models in 1D, with More-or-Less Intermittency.** (a) Exponents  $K(q)$  for a log-normal multifractal case and two monofractal cases; the onset of divergence for  $q > 4$  is indicated for the 1D log-normal model with  $C_1 = 0.25$ . (b) Four cascades with  $\langle \epsilon \rangle = 1$  that illustrate an increasing degree of intermittency:  $C_1 = 0.0, 0.05, 0.18$ , and  $0.72$ . Models with  $C_1 > 1$  would be “degenerate” in 1D, meaning almost empty in most realizations but a huge peak would occur now and then.



(a)



(b)

<sup>1</sup>The characteristic function for normal deviates is  $\langle \exp[itN(\mu, \sigma)] \rangle = \exp[i\mu t - t^2(\sigma^2/2)]$ , where we set  $t = q/i$ .

<sup>2</sup>Schertzer and Lovejoy (1987) propose a class of log-Lévy cascade models parameterized by  $\alpha \in [0, 2]$ . They run the gamut between beta-models in Eq. (A21) and log-normal models in Eq. (A29): if  $\alpha \neq 1$ ,  $K(q) = C_1(q^\alpha - q)/(\alpha - 1)$ , and  $K(q) = C_1 q \ln q$  for  $\alpha = 1$ .

#### A.3.4 The Microcanonical Log-Binomial “ $p$ -model”

Meneveau and Sreenivasan’s (1987b) one-dimensional “ $p$ -model” is a microcanonical alternative to the canonical log-normal model. Here  $\lambda = 2$  and the weights are

$$W_i = 1 \pm (1-2p) = \begin{cases} W_- = 2p & (\text{Prob} = 1/2) \\ W_+ = 2-2p & (\text{Prob} = 1/2) \end{cases}, \quad 0 < p < 1/2 \quad (i = 0, \dots, n-1) \quad (\text{A31a})$$

in one sub-eddy, and

$$W'_i = 2 - W_i = \begin{cases} W_+ \\ W_- \end{cases} \quad (\text{A31b})$$

in the other, to conserve the total measure (cf. Fig. A6). The microcanonical nature of this model means that every sequence of  $\pm$ ’s yields the same binomially distributed values for  $\varepsilon_n(x)$ ,  $W_-^m W_+^{n-m}$  ( $m = 0, \dots, n$ ) with probability  $\binom{n}{m}/2^n$ ; only their order of occurrence changes. Such predictability—exact ergodicity in the sense of singular measures—in a calibration procedure for singularity analysis is obviously desirable.

The choice of weights in Eq. (A31a) yields

$$K(q) = \log_2[W_-^{q/2} + W_+^{q/2}] = \log_2[(2p)^{q/2} + (2-2p)^{q/2}] - 1, \quad (\text{A32})$$

for<sup>1</sup>  $q \in \mathbb{R}$ . Exponents of special interest are

- the information dimension:  $D_1 = 1 - C_1 = -[p \log_2 p + (1-p) \log_2 (1-p)]$  from  $q = 1$ ; and
- the spectral exponent:  $\beta_\varepsilon = 1 - \log_2[1 + (1-2p)^2]$  from  $q = 2$ .

The latter is plotted as a function of  $p$  in Fig. A7a where the inset illustrates a typical realization. The sample in question was generated using  $p = 0.35$ , as suggested by Meneveau and Sreenivasan’s (1987a) measurements of the dissipation field in turbulent flows; its more important characteristics are  $D_1 \approx 0.65$  ( $C_1 \approx 0.35$ ), and  $\beta \approx 0.88$ . The two limiting cases are familiar:  $p \rightarrow 0^+$  yields randomly placed  $\delta$ -functions,  $p \rightarrow 1/2^-$  (or  $\sigma \rightarrow 0^+$  for the log-normal case) weak  $1/f$ -type fluctuations from the unitary mean value.

#### A.3.5 The “ $p(3)$ -model”, A 2D Generalization of the “ $p$ -model”

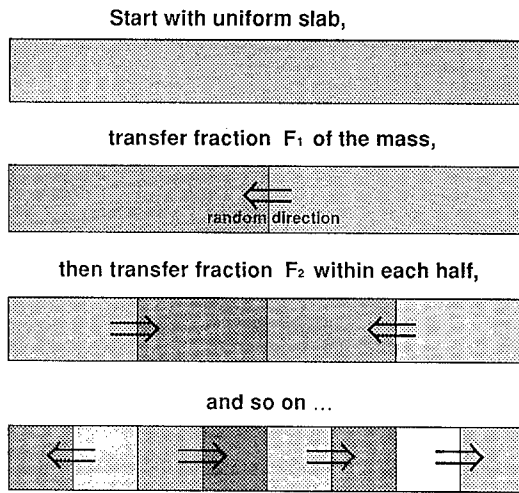
The most general microcanonical cascade with  $\lambda = 2$  in  $d = 2$  calls for three parameters that we will denote

$$0 \leq p_1, p_2, p_3 \leq 1/2. \quad (\text{A33})$$

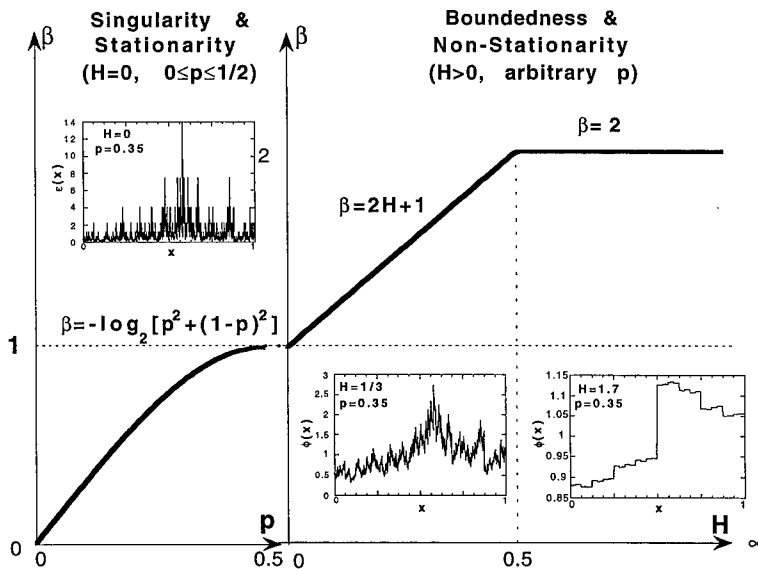
They will be used to shift mass in the same way as in the  $p$ -model but in both horizontal (E–W) and vertical (N–S) directions as well as between the two diagonals:

$$\left\{ \begin{array}{ll} \text{E} \leftrightarrow \text{W}: & 1 \pm (1-2p_x) \\ \text{N} \leftrightarrow \text{S}: & 1 \pm (1-2p_y) \\ \text{NE/SW} \leftrightarrow \text{NW/SE}: & 1 \pm (1-2p_d) \end{array} \right\}, \quad (\text{A34})$$

<sup>1</sup>We have  $K(q) \approx q \log_2 W_\pm - 1$  for  $q \rightarrow \pm\infty$ . From Eq. (A23b), and knowing that  $W_\pm = 2-2p < 2$ , the large  $q$  limit tells us that all moments converge for microcanonical models. Not only they converge but (by construction) their estimates for every realization are identical, even for a single cascade step.



**Figure A6:** Genesis of a Microcanonical Cascade in 1D. Three steps are illustrated. The total mass is conserved at each step, implying anti-correlated multiplicative weights in each sub-cell:  $W_i + W_i' = 2$ , hence  $W_i = 1 \pm F_i$  and  $W_i' = 1 \mp F_i$ .



**Figure A7:** Spectral Exponents of Singular and Bounded Cascades. (a) Stationary but intermittent case  $H = 0$ , reverting to Meneveau and Sreenivasan's (1987) " $p$ -model" for the kinetic energy dissipation rate that occurs at the Kolmogorov scale in fully developed turbulent flows; this model is unbounded ("singular") in the limit of many cascade steps. (b) Nonstationary  $H > 0$  generalization proposed by Cahalan *et al.* (1990, 1994) to model cloud optical depth variability in stratocumulus decks that are observed to have power-law spectra with  $\beta \approx 5/3$ .

where  $\{p_x, p_y, p_d\}$  is one of the  $3! = 6$  permutations of  $\{p_1, p_2, p_3\}$ , with equal probability and random  $\pm$ . After the three transfers, the individual combined weights are

$$W = \prod_{\theta \in \{x,y,d\}} W_\theta = \prod_{k=1}^3 [1 + s_k(1-2p_k)], \quad (\text{A35})$$

where  $\{s_k; k=1,2,3\} = \{s_1, s_2, s_3\} = \{\pm, \pm, \pm\}$ , one of the 8 equally probable combinations. These weights fall between

$$W_\pm = \min\{W\} = \prod_{k=1}^3 [1 \pm (1-2p_k)] \quad (\text{A36})$$

For this choice of  $W$ 's,

$$\langle W^q \rangle = \frac{1}{8} \sum_{\{\pm, \pm, \pm\}}^{8 \text{ combos}} \prod_{k=1}^3 [1 \pm (1-2p_k)]^q, \quad (\text{A37})$$

for  $q \in \Re$ , hence  $K(q)$  from Eq. (A23a). If  $p_1 = p < 1/2$  and  $p_2 = p_3 = 1/2$ , we have only two equally probable weights  $W_\pm$ , and Eq. (A32) is retrieved.

#### A.4 Going from Stationary Measures to Nonstationary Multifractal Functions

It is important to have the option of multiscaling as well as monoscaling (section A.2) in the realm of nonstationary functions with stationary increments. We therefore need algorithms that generate fields where structure-function analysis yields a nonlinear  $\zeta(q)$ , equivalently, a non-constant  $H(q)$ . There are many well-documented methods for generating stationary multifractal measures using multiplicative cascades; in contrast, the literature on specific ways of generating nonstationary functions with multiscaling structure functions is relatively small. We present two procedures here (Schertzer and Lovejoy, 1987; Cahalan *et al.*, 1990) and refer to Viscek and Barabási (1991), Arnéodo *et al.* (1992), Benzi *et al.* (1993), and Sykes *et al.* (1996) for the few others we are aware of.

##### A.4.1 "Bounded" Cascades

One route from multiplicative cascades to nonstationary multiscaling processes was charted by Cahalan *et al.* (1990) for the purposes of modeling the internal structure of marine stratocumulus. Their model builds on Meneveau and Sreenivasan's (1987)  $p$ -model, calling for a power-law decay of the variance of the weights as the multiplicative cascade proceeds to smaller and smaller scales,  $r_i = L/2^i$  ( $i = 0, \dots, n$ ). Explicitly, we take

$$W_i = 1 \pm (1-2p)(r_{i-1}/L)^{-H} \text{ (equal Prob } \pm), 0 < p \leq 1/2, 0 \leq H \leq \infty, \quad (\text{A38a})$$

$$W'_i = 2 - W_i, \quad (\text{A38b})$$

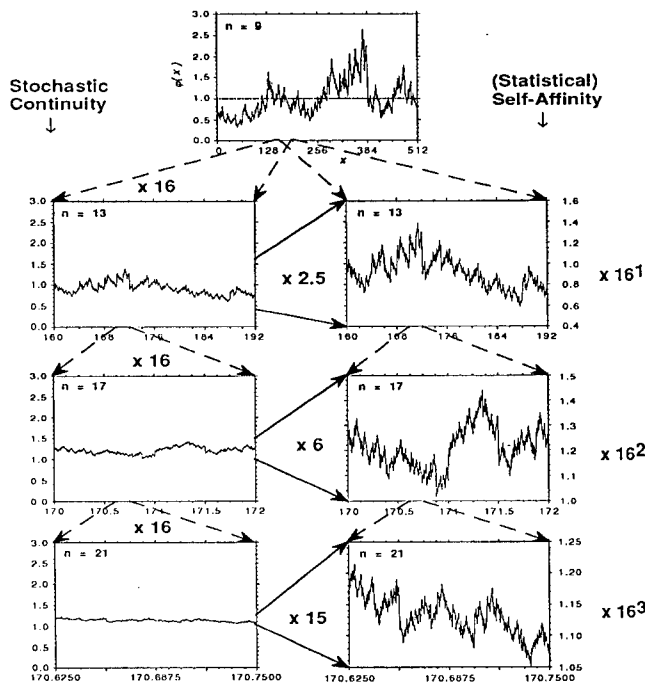
for  $i \geq 1$ . Figure A7b shows the effect of  $H$  on the spectral exponent:  $\beta = \min\{2H, 1\} + 1$  (Cahalan *et al.*, 1994); the inset shows examples for  $H = 1/3, 1.7$ .

Marshak *et al.* (1994) show that the resulting model,  $f_n(x)$ , is multifractal in the sense that higher-order structure functions:  $\langle |f_n(x+r) - f_n(x)|^q \rangle$  goes as  $r^q H(q)$  with

$$H(q) = \min\{H, 1/q\}. \quad (\text{A39})$$

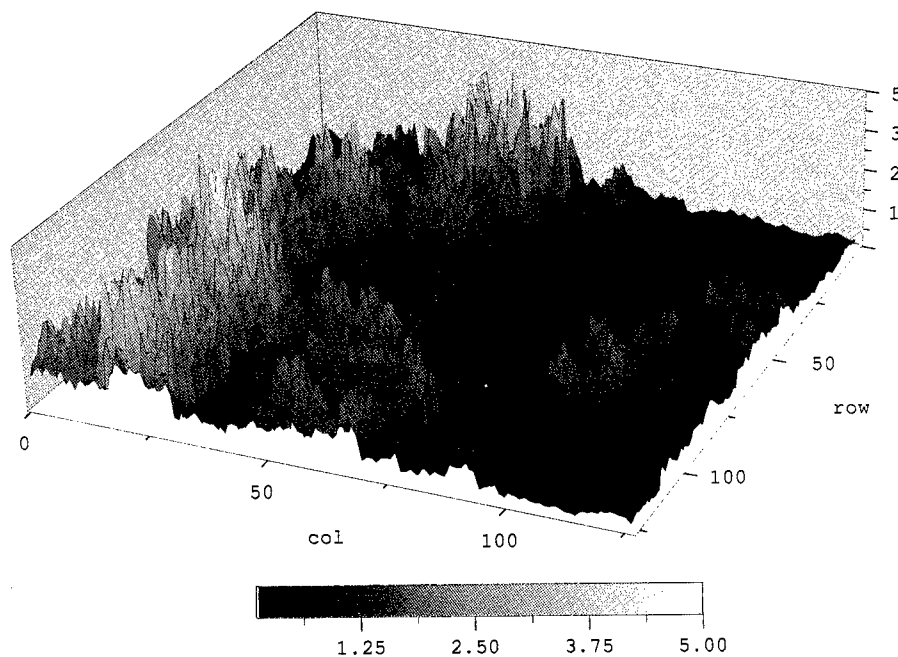
Interestingly,  $p$  does not appear in Eq. (A39) which determines the scaling of 2-point statistics of all orders; Cahalan *et al.* (1994) investigate its role in 1-point statistics, hence prefactors in Eq. (16) of the main text. Clearly,  $H$  controls the degree of nonstationarity. The  $p$ -model is retrieved in the “stationary” limit  $H = 0$ . At  $H \rightarrow \infty$ , we find the “most nonstationary” in this continuum of models, Heaviside steps of height  $2(1-2p)$  at  $x = L/2$ .

Figure A8 illustrates two important properties of bounded cascades which are shared by other multiscaling random functions (as well as fBm's, their monoscaling counterparts): stochastic continuity and self-affinity. Two identical sequences of successive horizontal zooms are illustrated; the difference is only that on the l.h.s. the vertical scale is held constant and on the r.h.s. it is rescaled by a given factor at each zoom. On the l.h.s. we see the amplitude of the increments decrease dramatically with scale (continuity property). However, on the r.h.s. we see that zooms onto different portions of the graph of  $f_n(x)$  are statistically similar: the graph is self-affine with fractal dimension  $D_{\text{graph}} = 2 - H_1 = 5/3$  in this  $H_1 = H = 1/3$  case.



**Figure A8:** *Stochastic Continuity and Self-Affinity of a 1D Bounded Cascade Model.* The sequence of zooms on the l.h.s. (without vertical rescaling) shows how smaller scales lead to smaller increments; this is called “stochastic continuity.” The r.h.s. sequence (with vertical rescaling) yields three graphs that are statistically indistinguishable from the original at the top; this property is “self-affinity.”

Figure A9 shows an example of a 2D generalization of the above bounded cascade model starting with the “ $p(3)$ -model.” The factors  $1-2p_k$  ( $k = 1, 2, 3$ ) in Eqs. (A34–36) are multiplied by the variance moderating factor  $(r_{i-1}/L)^{-H}$ ,  $i = 1, \dots, n$ , as in Eq. (38a). This type of model was used by Cahalan (1994), Marshak *et al.* (1995a) and Davis *et al.* (1996b) in radiative transfer computations.



**Figure A9:** *Perspective View of a 2D Bounded Cascade Model.* The numerical values of parameters are  $H = 1/3$  and  $p_k = 0.42 + (k-1) \times 0.03$  in a 7-step process.

#### A.4.2 Random Devil's Staircases

The simplest way of obtaining a function from a multifractal measure is to take its indefinite integral:

$$f_{\varepsilon}(x) = \int_0^x \varepsilon(x') dx', \quad 0 \leq x \leq L. \quad (\text{A40})$$

Since  $\varepsilon(x) \geq 0$ , this random "Devil's staircase" (Mandelbrot 1983) is a non-decreasing function, as illustrated by the typical realization based on a log-normal cascade in Fig. A10a. The increments of (A40) are easily computed, using the definition of a coarse-grained measure in Eq. (14) in the main body:

$$f_{\varepsilon}(x+r) - f_{\varepsilon}(x) = \int_x^{x+r} \varepsilon(x') dx' = r\varepsilon(r;x), \quad 0 \leq x \leq L-r, \quad 0 \leq r \leq L. \quad (\text{A41})$$

The  $q$ th order moment of this increment is directly related to that of  $\varepsilon(r;x)$ :

$$\langle |f_{\varepsilon}(x+r) - f_{\varepsilon}(x)|^q \rangle = r^q \langle \varepsilon(r;x)^q \rangle, \quad q < q_D. \quad (\text{A42})$$

Using the definitions in Eqs. (16) and (20) and taking logs, we find  $\zeta(q) = q - K(q)$ ; equivalently,

$$H(q) = 1 - \frac{K(q)}{q}, \quad q < q_D. \quad (\text{A43})$$

In particular, integrals of randomly placed  $\delta$ -functions, with  $K(q) = q-1$  ( $q > 0$ ), yield randomly placed Heaviside steps which are multifractal functions with a broad range of Hölder exponents:  $H(q) = 1/q$  ( $q > 0$ ). The limit  $H \rightarrow \infty$  in Eq. (A39) confirms this result, that Marshak *et al.* (1996) derive from first principles as well.

#### A.4.3 Fractionally Integrated Singular Cascades

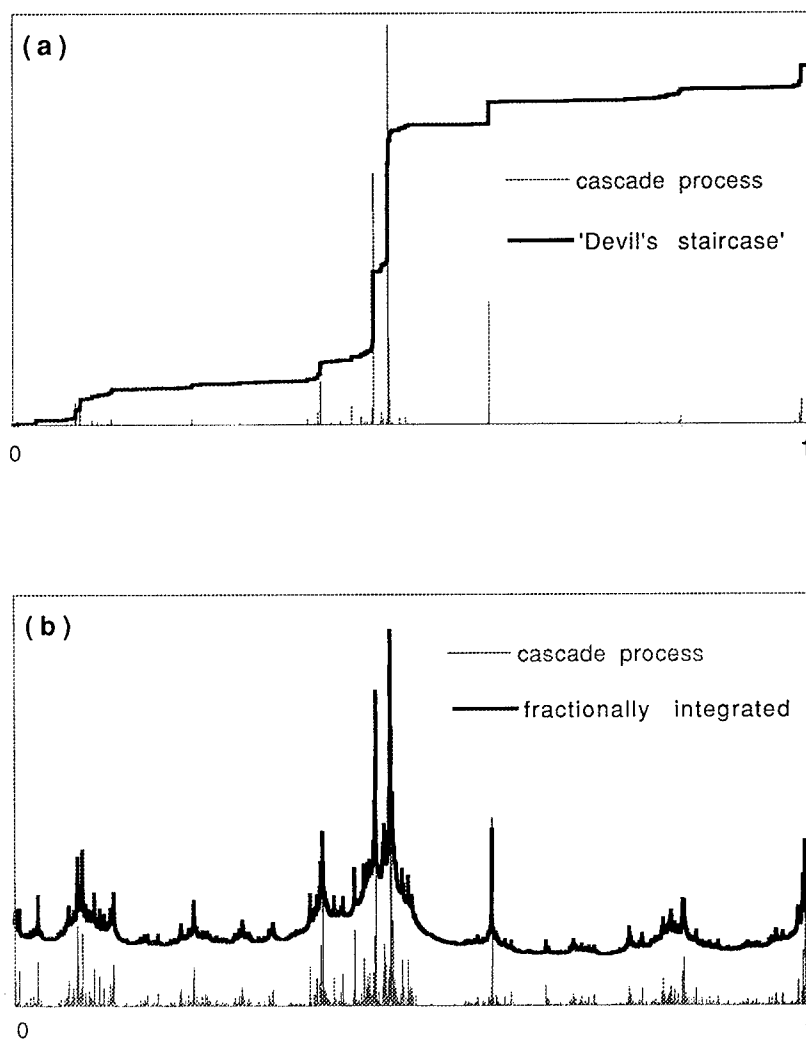
To simulate the scale-invariant spatial properties observed in cloud fields, Schertzer and Lovejoy (1987) generalize the Devil's staircase concept. As a means of introducing the continuity that necessarily comes with nonstationarity, they simply use fractional integration instead of its standard counterpart. This yields what shall call a Fractionally Integrated Singular Cascade (FISC). We use here the same form of fractional integration as in §A.2.3, low-pass power-law filtering in Fourier space. Consider a 1D grid of size  $M_n = 2^n$  and unitary outer scale,

$$x_m = j\ell_n, \quad m = 0, \dots, M_n - 1 \quad (L = M_n\ell_n = 1), \quad (\text{A44})$$

where we construct an  $n$ -step cascade process with branching ratio  $\lambda = 2$ . This field is generated in physical space and its Fourier representation is computed numerically:

$$\begin{cases} \varepsilon_n(x_j), & j = 0, \dots, 2^n - 1 \\ \tilde{\varepsilon}_n(k), & k = 0, \pm 1, \dots, \pm 2^{n-1}. \end{cases} \quad (\text{A45})$$





**Figure A10:** Hybrid Multiplicative/Additive Models. (a) Random Devil's Staircase: both the measure and its integral are illustrated. (b) Same as in panel (a) but for a fractional integration of order  $1/3$ .

The measure  $\varepsilon(x)$  is then “smoothed” into a function using

$$\begin{cases} f_n(x) = \varepsilon_n(x) * |x|^{-(1-H^*)} \\ \tilde{f}_n(k) = A(H^*) \tilde{\varepsilon}_n(k) \times |k|^{-H^*}, \end{cases} \quad 0 < H^* < 1, \quad (\text{A46a})$$

where (e.g., Gradstein and Ryzhik 1980)

$$A(H^*) = \sqrt{\frac{2}{\pi}} \cos\left(\frac{\pi}{2} H^*\right) \Gamma(H^*). \quad (\text{A46b})$$

Figure A10b shows the outcome for a log-normal cascade similar to that in Fig. A10a using a fractional order of integration  $H^* = 1/3$ . The exponent  $H^*$  of the power-law filter is the nonstationarity parameter of the model, given directly by the spectral exponent we want for  $f(x)$ :

$$H^* = \frac{\beta_f - \beta_\varepsilon}{2}, \quad (\text{A47})$$

given that of  $\varepsilon(x)$ ,

$$\beta_\varepsilon = 1 - K(2), \quad (\text{A48})$$

where  $K(2)$  defines the scaling of  $\langle \varepsilon(r;x)^2 \rangle \sim r^{-K(2)}$ . As for fBm using the Fourier construction (§A.2.3), generalization to 2D is straightforward.

We need to estimate the scaling exponents for this model for  $q \neq 2$  (at  $q = 2$ , the scaling is determined exactly by the construction algorithm). For this, we tentatively interpret Eq. (46a) as

$$|f(x+r) - f(x)| \sim \varepsilon(r;x) r^{H^*}; \quad (\text{A49})$$

taking  $q$ th powers, and averaging yields  $\zeta(q) = qH^* - K(q)$ , given the exponent definitions in main text's Eqs. (16) and (20). This  $\zeta(q) \leftrightarrow K(q)$  connection is a special case of Eq. (26) with  $a = 1/H^*$  and  $b = 1$ , leading to

$$H(q) = H^* - \frac{K(q)}{q}, \quad q < q_D. \quad (\text{A50})$$

So this generalizes Eq. (A43) for Devil's staircases where  $H^* = 1$ .

However, the general (event-wise) applicability of (A49), hence (A50) for all values of  $q$ , to FISCs is questionable. However, Eq. (A50) is guaranteed by construction to work for  $q = 0$  and  $q = 2$ , between these two values it provides at least a good approximation to numerically obtained  $\zeta(q)$ 's. As  $q$  increases far beyond 2, we are effectively emphasizing ever larger events in  $\varepsilon(r;x)$  on the r.h.s. of (A49). It is easy to imagine situations where the increment  $f(x+r) - f(x)$  on the l.h.s. of (A49) is small, although the underlying measure  $\varepsilon(r;x)$  is large. For instance, take  $x+r$  and  $x$  on either side of a strong spike in Fig. A10b. Consequently, the agreement between the statistics independently determined on either side of Eq. (A49) deteriorates.<sup>1</sup>

<sup>1</sup>Davis *et al.* (1993, 1994b) describe a more general approach for characterizing the correlations between some nonstationary process  $f(x)$ , artificial or real, and the associated  $\varepsilon(x)$  field.

**Table A1 (begin): Scale-Invariant Processes.** 2nd-order statistics and associated stationarity and continuity properties.

Model	Sect.	Figure	Domains?		Parameters?		Stationarity?		
			$x \in [0, L]^d$ $w, f$ or $\varepsilon$		primary	secondary	... per se <sup>a</sup> ( $\beta < 1$ )	... of incs <sup>b</sup> ( $\beta < 3$ )	
exponents:			$d$				$\beta$		
<b>I. Discrete model for <math>w(x) = f'(x)</math>, without correlations:</b>									
a Bernoulli w.n.	A.1.1		1	$w = \pm s$	$s > 0$		0	✓	✓
b Bm on a grid	A.1.1		1	$f/s \in \mathbb{Z}$	$s > 0$		2		✓
<b>II. Gaussian model for <math>w(x) = f'(x)</math>, without correlations:</b>									
a Gaussian w.n.	A.1.1	A1a	1,2	$w \in \mathcal{R}$	$\sigma > 0$	$\mu$	0	✓	✓
b off-grid Bm	A.1.1	A1b	1	$f \in \mathcal{R}$	$\sigma > 0$	$f_0$	2		✓
<b>III. Non-Gaussian model for <math>w(x) = f'(x)</math>, without correlations:</b>									
a Lévy w.n.	A.1.2	A2a	1,2	$w \in \mathcal{R}$	$0 < \alpha < 2$	$a, c$	0	✓	✓
b Lévy-flight	A.1.2	A2b	1	$f \in \mathcal{R}$	$0 < \alpha < 2$	$c, f_0$	2		✓
<b>IV. Gaussian models for <math>f'(x)</math> and <math>f(x)</math>, with correlations:</b>									
a SGSN <sup>+</sup>	A.2.4		1,2	$f' \in \mathcal{R}$	$-1 < H_2 < 0$	$\sigma$	$1-2 H_2 $	✓	✓
b fBm	A.2	A3	1,2	$f \in \mathcal{R}$	$0 < H_2 < 1$	$\sigma$	$2H_2+1$		✓
<b>V. Cascade models <math>\varepsilon(x)</math> and <math>f(x)</math>, with microcanonical conservation:</b>									
a p-model	A.3.4	A6-7	1,	$\varepsilon > 0$	$0 < p < 1/2$		$1-\log_2[1+(1-2p)^2]$	✓	✓
3p-model	A.3.5		2		$0 < p_i < 1/2$ ( $i = 1, 2, 3$ )		$1-\log_2\langle W^2 \rangle$ , with $q=2$ in (A33).		
b Bounded Cascades	A.4.1	A6-8 A9	1, 2	$f_- \leq f \leq f_+$ ( $0 < f_- < f_+ < \infty$ )	$H > 0$	$p,$ $p_i$ ( $i=1,2,3$ )	$\min\{2H, 1\}+1$		✓
<b>VI. Cascade models for <math>\varepsilon(x)</math> and a "hybrid" one for <math>f(x)</math>, with canonical conservation:</b>									
a Beta-models	A.3.1	A4a, A5a	1,2	$\varepsilon \geq 0$	$0 < D_f < d$		$1-(d-D_f)$	✓	✓
a' lognormal casc.	A.3.3	A4b, A5	1,2	$\varepsilon > 0$	$\sigma_{\ln W} > 0$		$1-\sigma_{\ln W}^2/\ln 2$	✓	✓
b FISC*	A.4.3	A10b	1,2	$f > 0$	$H^* > (1-\beta_\varepsilon)/2$ $p, D_0$ , or $\sigma_{\ln W}$		$2H^*+\beta_\varepsilon$		✓

<sup>a</sup>Models designated with an "a" are stationary in the "broad" sense, where the autocorrelation function depends only on the separation:  $\langle w(x+r)w(x) \rangle \propto \delta(r)$  (I–III);  $\langle f'(x+r)f'(x) \rangle \propto r^{2H_2}$  (IV, with  $-1/2 < H_2 < 0$ );  $\langle \varepsilon(x+r)\varepsilon(x) \rangle \propto r^{-K(2)}$  (V–VI, with  $0 < K(2) = 1-\beta < 1$ ). These models are also stochastically discontinuous since, e.g.,  $\langle [w(x+r)-w(x)]^2 \rangle = 0$  for  $r \geq 0$ .

<sup>b</sup>The associated models designated with a "b" are (broad-sense) nonstationary but stochastically continuous in the sense that, as  $r \rightarrow 0$ ,  $\langle [f(x+r)-f(x)]^2 \rangle \propto r^{\zeta(2)} \rightarrow 0$  with  $0 < \zeta(2) = \beta-1 < 2$ . In categories I–III, one can go from model "a" to model "b" in  $d=1$  simply by taking a running sum; they are "additive" models.

\*Scaling Gaussian Stationary Noise.

\*Fractionally Integrated Singular Cascade.

**Table A1 (end): Scale-Invariant Processes.** Multifractal statistics and associated ergodicity properties.

Multi-scaling?		Inter-mittency?		Bi-fractal properties?			Ergodicity Properties?
$\zeta(q)$ & ... $K(q)$		$(K(q) \neq 0)$ for $w, f$ or $\epsilon$ & ... for $ \nabla $ 's		(position in $q = 1$ plane)			(remarks)
				$H_1(w, f \text{ or } \epsilon)$	$C_1(w, f \text{ or } \epsilon)$	$C_1( \nabla $ 's)	
no	N.A.	N.A.	no	0	N.A. (negative values)	0	trivial ergodicity
no	N.A.	N.A.	no	1/2	N.A. ( " " )	0	large fluctuations in local 1-pt. statistics and realization-to-realization differences.
no	N.A.	N.A.	no	0	N.A. ( " " )	0	$10^3$ events are generally enough to sample up to $\approx 3\sigma$ 's (cf. "3 $\sigma$ " rule).
no	N.A.	N.A.	no	1/2	N.A. ( " " )	0	cf. Ib
no	N.A.	N.A.	no	0	N.A. ( " " )	0	moments of order $q \geq \alpha$ are divergent.
yes	N.A.	N.A.	no	$\min\{1/\alpha, 1\}$	N.A. ( " " )	0	same as Ib and IIb, but worse due to divergences.
no	N.A.	N.A.	no	0	N.A. ( " " )	0	cf. IIa
no	N.A.	N.A.	no	$H_2$	N.A. ( " " )	0	cf. IIb
no	yes	yes	yes	0	$p\log_2 p + (1-p)\log_2 (1-p) + 1$  $(d/dq)\log_2 \langle W^q \rangle _{q=1} = \langle W\log_2 W \rangle$ , from (A33).	$C_1(\epsilon)^\#$	$\langle \epsilon(r, x) \rangle$ distribution is independent of the realization.  $\langle \epsilon(L, 0) \rangle$ is independent of realization.
yes	no	no	N.A. <sup>†</sup>	$\min\{H, 1\}$	0	N.A. <sup>†</sup>	although on a bounded domain, between $f_{\pm} = \prod_{i=0}^{\infty} [1 \pm (1-2p)/2^i H]$ , the pdf has lognormal-like skewness and the associated sampling problems; increments are also broadly distributed.
no	no	yes	yes	0	$d - D_f$	$C_1(\epsilon)^\#$	no divergence in $\langle \epsilon(r, x)^q \rangle$ .
no	yes	yes	yes	0	$\sigma_{\ln W}^2/2\ln 2$	$C_1(\epsilon)^\#$	$\langle \epsilon(r, x)^q \rangle$ diverges for $q \geq d/C_1$ .
yes	no	no	yes	$H^*$	0	$C_1 \epsilon^\S$	$K(q)$ for $ \nabla $ 's: cf. corresponding $\epsilon$ -model.

<sup>†</sup>In standard singularity analysis using next-neighbor absolute differences, spurious scale-breaks occur due to the binary-tree construction of this model.

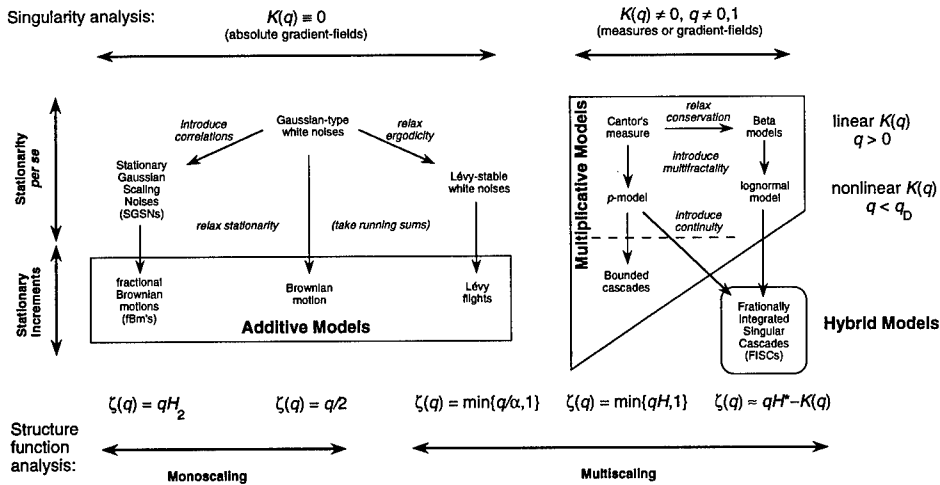
<sup>#</sup>Taking absolute gradients of a singular cascade model only emphasizes the spikes, and the same singularity spectrum is found (Lavallée *et al.*, 1993).

<sup>§</sup>Absolute gradients of a FISC have the same singularity properties as the original cascade (Lavallée *et al.*, *ibid.*).

### A.5 Summary

We have described specific algorithms for constructing mono- and multiscaling stationary measures amenable to singularity analysis (Section 4.4 in main text), as well as mono- and multiscaling nonstationary functions with stationary increments amenable to structure function analysis (Section 4.2 in main body). In each case, we have specified the dependency of the appropriate scaling exponents on the parameters of the model. One example in two spatial dimensions is described explicitly and, in all others, generalization from 1D to 2D is straightforward. This collection of models is comprehensive enough to calibrate and study the sensitivity of any standard multifractal data analysis procedure using arbitrary amounts of synthetic data with controllable statistical properties. In particular, convergence rates of spatial averages to their ensemble counterparts can be investigated since the models have variable degrees of ergodicity. Table A1 organizes the models by category, lists their parameters and summarizes their properties, while Fig. A11 displays their inter-relations graphically.

Validation of analysis procedures is only one application for stochastic modeling. In our specific area of research, the effects of internal cloud structure on radiative properties, we have used multifractal models as artificial clouds-in-a-computer with controllable properties. Extensive numerical experimentation, using Monte Carlo and other radiative transfer techniques, has lead us to new insight into the ways clouds affect the Earth's radiative budget (Cahalan 1994) and ways of retrieving cloud properties from satellite (Marshak *et al.* 1995*a,b*) and lidar (Davis *et al.* 1996*b*) data.



**Figure A11: Classification of Scale-Invariant Models.** The formulas for the  $q$ th-order structure function exponents refer to the models in the last row that are nonstationary but with stationary increments.

## Acknowledgments

This work was supported by the Environmental Sciences Division of U.S. Department of Energy (under grant DE-A105-90ER61069 to NASA's Goddard Space Flight Center) as part of the Atmospheric Radiation Measurement (ARM) program. We thank A. Arnéodo, T. Bell, S. Lovejoy, Ch. Meneveau, D. Schertzer, and T. Warn for helpful discussions.

## References

- Arnéodo, A., Grasseau, G., and Holschneider, M., "Wavelet transform of multifractals," *Phys. Rev. Lett.* **61** (1988) 2281–2284.
- Arnéodo, A., J. F. Muzy, and E. Bacry, "Wavelet analysis of fractal signals. Applications to fully developed turbulence data," in *Eddy Structure Identification in Free Turbulent Shear Flows*, eds. J.P. Bonnet and M.N. Glauser (1992), p. 153.
- Aurell, E., U. Frisch, J. Lutsko, and M. Vergassola, "On the multifractal properties of the energy dissipation derived from turbulence data," *J. Fluid Mech.* **238** (1992) 467–486.
- Benzi, R., L. Biferale, A. Crisanti, G. Paladin, M. Vergassola, and A. Vulpani, "A random process for the construction of multi-affine fields," *Physica D* **65** (1993) 352–358.
- Bertozzi, A. L., and A. B. Chhabra, "Cancellation exponents and fractal scaling," *Phys. Rev. E* **49** (1994) 4716–4719.
- Cahalan, R. F., M. Nestler, W. Ridgway, W. J. Wiscombe, and T. L. Bell, "Marine stratocumulus spatial structure," in *Proceedings of the 4th International Meeting on Statistical Climatology*, ed. J. Sansom (New Zealand Meteorological Service, Wellington, 1990), pp. 28–32.
- Cahalan, R. F., W. Ridgway, W. J. Wiscombe, T. L. Bell, and J. B. Snider, "The albedo of fractal stratocumulus clouds," *J. Atmos. Sci.* **51** (1994) 2434–2455.
- Cahalan, R. F., "Bounded cascade clouds: albedo and effective thickness," *Nonlin. Proc. in Geophysics* **1** (1994) 156–167.
- Christakos, G., *Random Fields in Earth Sciences* (Academic Press, San Diego, 1992).
- Coakley, J. A. (Jr.), and F. R. Bretherton, "Cloud cover from high-resolution scanner data: Detecting and allowing for partially filled fields of view," *J. Geophys. Res.* **87** (1982) 4917–4932.
- Corrsin, S., "On the spectrum of isotropic temperature fluctuations in isotropic turbulence," *J. Appl. Phys.* **22** (1951) 469–473.
- Davis, A., A. Marshak, and W. Wiscombe, "Bi-multifractal analysis and multi-affine modeling of nonstationary geophysical processes – application to turbulence and clouds," *Fractals* **1** (1993) 560–567.
- Davis, A., A. Marshak, W. Wiscombe, and R. Cahalan, "Multifractal characterizations of nonstationarity and intermittency in geophysical fields, observed, retrieved or simulated," *J. Geophys. Res.* **99** (1994a) 8055–8072.

- Davis, A., A. Marshak, and W. Wiscombe, "Wavelet-based multifractal analysis of nonstationary and/or intermittent geophysical signals," in *Wavelets in Geophysics*, eds. E. Foufoula-Georgiou and P. Kumar (Academic Press, San Diego, 1994b), pp. 249–298.
- Davis, A., A. Marshak, W. Wiscombe, and R. Cahalan, "Scale-invariance in liquid water distributions in marine stratocumulus – Part I: Spectral properties and stationarity issues," *J. Atmos. Sci.* (1996a, in press).
- Davis, A., A. Marshak, R. Cahalan, and W. Wiscombe, "Horizontal radiative fluxes in stratocumulus and the landsat scale-break," *J. Atmos. Sci.* (1996b, accepted).
- Davis, A., and A. Marshak, "Bi-Fractal Analysis of Atmospheric Signals Illustrated with Cloud Data," in *Proceedings of the 1995 Battelle Atmospheric Conference* (1996, in press).
- Eneva, M., "Fractal description of microseismic activity," *Nonlin. Proc. in Geophysics* **1** (1994) 65–76.
- Evertsz, C. J. G., and B. B. Mandelbrot, "Multifractal measures," in *Chaos and Fractals, New Frontiers of Science*, eds. H.-O. Peitgen, H. Jürgens and D. Saupe (Springer-Verlag, New York, 1992), pp. 921–953.
- Falconer, K. J., *Fractal Geometry - Mathematical Foundations and Applications*, pp. xxii+288, J. Wiley, New York, 1990.
- Feigenbaum, M. J., "Some characterizations of strange sets," *J. Stat. Phys.* **46** (1987) 919–924.
- Feller, W., *An Introduction to Probability Theory and its Applications*, vol. 2, (Wiley, New York, 1971).
- Frisch, U., P. L. Sulem, and M. Nelkin, "A simple dynamical model of intermittent fully developed turbulence," *J. Fluid Mech.* **87** (1978) 719–724.
- Frisch, U., "From global scaling, à la Kolmogorov, to local multifractality in fully developed turbulence," *Proc. R. Soc. Lond. A* **434** (1991) 89–99.
- Gerber, H., B. G. Arends, and A. S. Ackerman, "New microphysics sensor for aircraft use," *Atmos. Res.* **31** (1994) 235–252.
- Gradshteyn, I. S., and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Academic Press, San Diego, 1980).
- Grassberger, P., "Generalized dimensions of strange attractors," *Phys. Rev. Lett.* **97** (1983) 227–.
- Grivet-Talocia, S., "On the scale dependence of fractal dimension for band-limited  $1/f^\alpha$  noise," *Physics Lett. A* **200** (1995) 264–276.
- Gupta, V. K., and E. C. Waymire, "A statistical analysis of mesoscale rainfall as a random cascade," *J. Appl. Meteor.* **32** (1993) 251–267.
- Halsey, T. C., M. H. Jensen, L. P. Kadanoff, I. Procaccia, and B. Shraiman, "Fractal measures and their singularities – The characterization of strange sets," *Phys. Rev. A* **33** (1986) 1141–1151.
- Hentschel, H. G. E., and I. Procaccia, "The infinite number of generalized dimensions of fractals and strange attractors," *Physica D* **8** (1983) 435–444.
- Jensen, J. L. W. V., "Sur les fonctions convexes et les inégalités entre les valeurs moyennes," *Acta Math.* **30** (1906) 789–806.

- Kahane, J. P., and J. Perrière, "Sur certaines martingales de Benoit Mandelbrot," *Adv. in Math.* **22** (1976) 131-145.
- Kolmogorov, A. N., "Local structure of turbulence in an incompressible liquid for very large reynolds numbers," *Dokl. Akad. Nauk SSSR* **30** (1941) 299-303.
- Kolmogorov, A. N., "A Refinement of Previous Hypothesis Concerning the Local Structure of Turbulence in Viscous Incompressible Fluid at High Reynolds Number," *J. Fluid Mech.* **13** (1962) 82-85.
- Lavallée, D., S. Lovejoy, D. Schertzer, and P. Ladoy, "Nonlinear variability and simulation of landscape topography," in *Fractals in Geography*, eds. L. De Cola and M. Lam (Kluwer, Boston, 1993) pp. 158-192.
- Mallat, S. G., "A theory for multiresolution signal decomposition: the wavelet transformation," *IEEE Trans. Pattern Anal. Mach. Intel.* **11** (1989) 674-693.
- Marshak A., A. Davis, R. Cahalan, and W. Wiscombe, "Bounded cascade models as nonstationary multifractals," *Phys. Rev. E* **49** (1994) 55-69.
- Marshak A., A. Davis, W. Wiscombe, and G. Titov, "The verisimilitude of the independent pixel approximation used in cloud remote sensing," *Remote Sens. Environ.* **52** (1995a) 71-78.
- Marshak A., A. Davis, and W. Wiscombe, "Radiation smoothing in fractal clouds." *J. Geophys. Res.* **100** (1995b) 26247-26261.
- Marshak, A., A. Davis, W. J. Wiscombe, and R. F. Cahalan, "Scale-invariance of liquid water distributions in marine stratocumulus - Part 2: Multifractal properties and intermittency issues," *J. Atmos. Sci.* (1996, submitted).
- Mandelbrot, B. B., "Intermittant turbulence and self-similar cascades: Divergence of high moments and dimension of the carrier," *J. Fluid Mech.* **62** (1974) 331-350.
- Mandelbrot, B. B., *The Fractal Geometry of Nature* (Freemann, New York, 1983).
- Mandelbrot, B. B., and J. W. van Ness, "Fractional Brownian motions, fractional noises and applications," *SIAM Review* **10** (1968) 422-437.
- Meneveau, C., and K. R. Sreenivasan, "The multifractal spectrum of the dissipation field in turbulent flows," *Nuclear Phys. B* **2** (1987a) 49-76.
- Meneveau, C., and K. R. Sreenivasan, "Simple multifractal cascade model for fully developed turbulence," *Phys. Review Lett.* **59** (1987b) 1424-1427.
- Meneveau, C., and K. R. Sreenivasan, "Multifractal nature of turbulent energy dissipation," *J. Fluid Mech.* **224** (1991) 429-.
- Monin, A. S., and A. M. Yaglom, *Statistical Fluid Mechanics*, vol. 2 (MIT Press, Boston, 1975).
- Muzy, J.-F., E. Bacry, and A. Arnéodo, "Multifractal formalism for fractal signals: The structure-function approach versus the wavelet-transform modulus-maxima method," *Phys. Rev. E* **47** (1993) 875-884.
- Muzy, J.-F., E. Bacry, and A. Arnéodo, "The multifractal formalism revisited with wavelets," *Int. J. of Bifurcation and Chaos*, **4** (1994), 245-302.
- Novikov, E. A., and R. Stewart, "Intermittancy of turbulence and spectrum of fluctuations in energy dissipation," *Izv. Akad. Nauk. SSSR, Ser. Geofiz.* **3** (1964) 408-412.



- Obukhov, A., "Structure of the temperature field in a turbulent flow," *Izv. Akad. Nauk SSSR, Ser. Geogr. i Geofiz.* **13** (1949) 55–69.
- Obukhov, A., "Some specific features of atmospheric turbulence," *J. Geophys. Res.* **67** (1962) 3011–3014.
- Ott, E., Y. Du, K.R. Sreenivasan, A. Juneja, and A.K. Suri, "Sign-singular measures: Fast magnetic dynamos, and high Reynolds-number fluid turbulence. *Phys. Rev. Lett.* **69** (1992) 2654–2657.
- Parisi, G., and U. Frisch, "A multifractal model of intermittency," in *Turbulence and Predictability in Geophysical Fluid Dynamics and Climate Dynamics*, eds. M. Ghil, R. Benzi, and G. Parisi (North-Holland, 1985), pp. 84–88.
- Pearson, C. E. (ed.), *Handbook of Applied Mathematics – Selected Results and Methods*, 2nd ed. (Van Nostrand Reinhold, New York, 1990).
- Peitgen, H.-O., and D. Saupe (eds.), *The Science of Fractal Images* (Springer-Verlag, 1988).
- Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in FORTRAN*, 2nd ed. (Cambridge University Press, New York, 1993).
- Schertzer, D., and S. Lovejoy, "Physical modeling and analysis of rain clouds by anisotropic scaling multiplicative processes. *J. Geophys. Res.* **92** (1987) 9693–9714.
- Schertzer D., and S. Lovejoy, "Hard and soft multifractal processes," *Physica A* **185** (1992) 187–194.
- Sreenivasan, K. R., "Fractals and multifractals in fluid turbulence," *Ann. Rev. Fluid. Mech.* **23** (1991) 539–600.
- Sykes, R. I., R. S. Gabruk, and D. S. Henn, "A fractal/multifractal representation of the small-scale structure in a turbulent plume," *J. Appl. Meteor.* (1996, in press).
- Vainshtein, S. I., K. R. Sreenivasan, R. T. Pierrehumbert, V. Kashyap, and A. Juneja, "Scaling exponents for turbulence and other random processes and their relationships with multifractal structure, *Phys. Rev. E* **50** (1994) 1823–1835.
- Viscek, T., and A.-L. Barabási, "Multi-affine model for the velocity distribution in fully turbulent flows," *J. Phys. A: Math. Gen.*, **24** (1991) L845–L851.
- Voss, R. F., "Fourier synthesis of Gaussian fractals: 1/f noises, landscapes, and flakes," in *Proc. Siggraph Conf.* (Detroit, 1983), pp. 1–21.
- Waymire, E., and V. J. Gupta, "The mathematical structure of rainfall representations, Parts 1–3," *Water Resour. Res.* **17** (1981) 1261–1294.
- Wiscombe, W., A. Davis, and A. Marshak, "Scale-invariance, nonstationarity and intermittency in the structure of cloud cover," in *Proceedings of the 4th Atmospheric Radiation Measurement (ARM) Science Team Meeting* (U.S. Dept. of Energy, Washington DC, 1995), pp. 11–14.

## SHORT-TIME AVERAGING OF TURBULENT SHEAR FLOW SIGNALS

A. K. GUPTA

*Department of Aerospace Engineering  
I. I. T. Kanpur U P 208 016 INDIA  
E-mail: ak Gupta@iitk.ernet.in*

Two detection schemes are described. The first, applied to a  $u'$  single hot-wire signal from the buffer region of a turbulent boundary layer considers the fluctuating cycle as a unit to calculate short time averages. In the second, VISA is applied to a  $u'$  single hot-wire signal from the separated region of a backward facing step to detect the flapping oscillations.

### 1. Introduction

The existence of coherent structures in turbulent shear flows has led to the search for new algorithms for their detection during the past three decades. The discovery of coherent structures was made possible by flow visualization techniques using hydrogen bubbles in a turbulent water boundary layer by Kline et al. (1967) and using shadowgraphs in a gaseous turbulent mixing layer by Brown and Roshko (1974). Subsequently, hot wire anemometry has been employed extensively to detect coherent structures using variable interval time averaging schemes by Gupta et al. (1971) (viscous sublayer streaks), and Blackwelder and Kaplan (1976) (burst frequency) amongst a large number of other investigations. During the last decade, coherent structures have been recognized in almost all type of turbulent shear flows. An example is the relatively simple geometry of backward facing step (BFS) flow. It would appear that coherent structures constitute an important and essential physical aspect of all stationary and nonstationary turbulent flows in which turbulence production is present.

The distinguishing features of coherent structures are: 1) their localization in space and time, and 2) their relatively high kinetic energy content thereby signifying bursts of turbulence or energetic events as opposed to quiescent flow or background turbulence. Conventional long time Reynolds averaging in the time domain and Fourier spectral analysis in the frequency domain have been unable to detect them; the former due to masking of energetic events by the length of averaging, and the latter due to lack of fixed periodicity of the energetic events.

Attempts have been made in the past to detect the coherent structures using the VITA (Variable Interval Time Averaging) algorithm in the time domain and the wavelet transform in the frequency domain. The VITA scheme of Blackwelder and Kaplan (1976) employs a fixed interval of short time averaging and has been used extensively to detect the burst frequency in a turbulent boundary layer. The wavelet transform, on the other hand, requires the specification of some generic shape of the wavelet (Farge, 1992) and therefore is similar to a pattern recognition algorithm in the frequency domain.

The quantification of relatively high kinetic energy content in the detection criterion continues to be a serious problem, although a threshold value approximately equal to the long-time variance of the turbulent signal has been employed by most investigators.

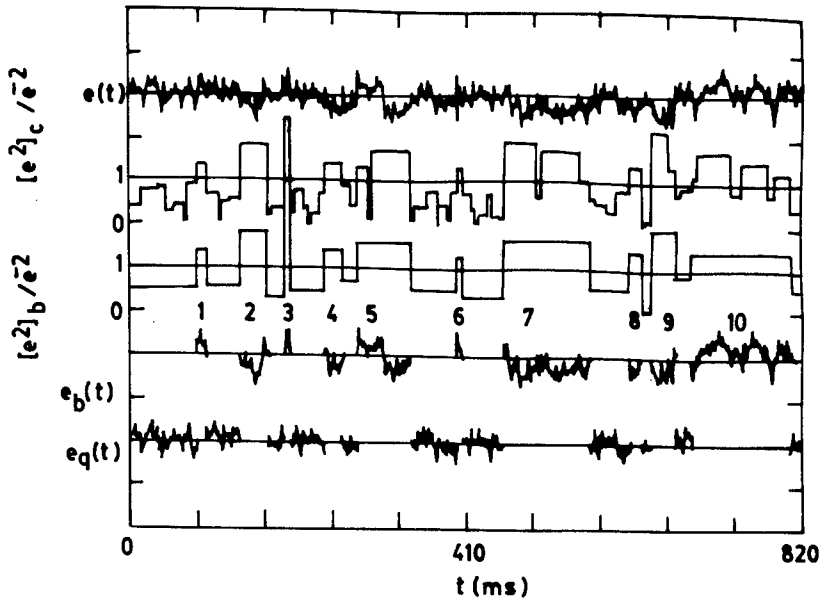
In the present paper, results obtained from a detection scheme applied to a  $u'$  single hot-wire signal from the buffer region ( $y^+ \approx 15$ ) of a turbulent boundary layer ( $R_\theta \approx 2300$ ) is described.  $y^+ = yu_* / \nu$  and  $R_\theta = U_\infty \theta / \nu$ , where  $y, U_\infty, u_*, \theta$ , and  $\nu$  are vertical distance from the wall, free stream wind speed, friction velocity, momentum thickness, and kinematic viscosity, respectively. In place of a fixed period for short-time averaging, the algorithm considers the fluctuating cycle of the turbulent flow signal about the conventional long-time mean value as a unit of averaging time. This results in variable averaging periods for the coherent structures. The duration of coherent structure, however, is decided by using long-time variance as the threshold value as most investigators have done in the past. Some of these results were presented by Gupta and Kaplan (1987).

The second detection scheme is variable interval spectral averaging (VISA), which was applied to a  $u'$  single hot-wire signal from the separated layer of a backward facing step flow to detect flapping oscillations using an HP 35660A Dynamic Signal Analyzer. This dynamic signal analyzer is able to vary the digitizing rate of the turbulent flow signal as well as to vary the number of blocks used for averaging the FFT spectra. Details of this investigation are given by Ananda (1993) and part of the results were presented by Ananda et al. (1992).

## 2. Statistical Characteristics of Bursts in a Turbulent Boundary Layer

Application of the first detection scheme to a block of single hot-wire data of 2048 samples (digitizing rate 2500 samples per second) is shown in Figure 1, which is a hard copy of the computer video monitor display.

The top trace shows the turbulent signal  $e(t)$  with the horizontal solid line denoting the long-time average  $\bar{E}$ . The second trace shows the plot of  $[e^2]_c / \bar{e^2}$  for each successive fluctuation cycle.  $[e^2]_c$  is the short-time variance computed for each fluctuating cycle, while  $\bar{e^2}$  is the long-time variance. The solid line  $[e^2]_c / \bar{e^2} = 1$  indicates the threshold value used to separate energetic burst events and quiescent events. It may be emphasized that the averaging time is the fluctuating cycle time and is therefore variable in the true sense by being different for each fluctuating cycle as shown in the second trace from above in Figure 1.



**Figure 1: Detection of Energetic and Quiescent Events;**  
 $y^+ \approx 15$ . Data Length = 2048 Samples.

The third trace  $[e^2]_b / \overline{e^2}$  shows the quiescent and energetic events alternating. These were obtained by merging the adjacent quiescent cycles into a quiescent event and adjacent energetic cycles into an energetic event.

Finally, the last two traces  $e_b(t)$  and  $e_q(t)$  show the splitting of the signal  $e(t)$  into an energetic burst signal and a quiescent signal, respectively. The numbers 1-10 on the trace  $e_b(t)$  identify the burst events 1-10.

Having delineated the events in a turbulent flow signal it was now possible to compute the detailed statistical properties of the energetic and quiescent events. An obvious quantity of interest is the mean bursting period  $\bar{T}_B$ . The bursting period is the time interval between two successive energetic events  $e_b(t)$ .  $\bar{T}_B$  is obtained from the histogram of the bursting periods. Figure 2 shows the histogram of bursting period carried out over the entire 50176 samples of a  $u'$  single hot-wire signal from the buffer region of a turbulent boundary layer developing in a low speed wind tunnel at the University of Southern California.

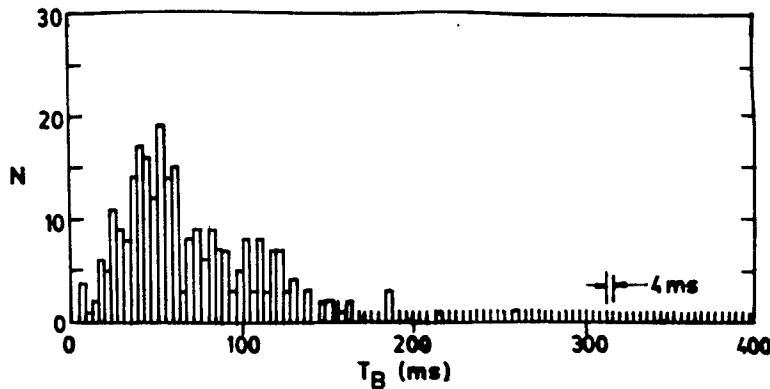


Figure 2: Histogram of Bursting Period;  
 $\bar{T}_B = 74 \text{ ms}$ ;  $\Sigma N = 271$

There were a total of 271 bursting events with a mean bursting period of  $\bar{T}_B = 74 \text{ ms}$  and standard deviation equal to  $31 \text{ ms}$ . In dimensionless form, the values of mean bursting period were  $\bar{T}_B U_\infty / \delta = 3.7$  and  $\bar{T}_B u_*^2 / \nu = 185$ , to be compared with the value of  $\bar{T}_B U_\infty / \delta = 5.0$  of Rao et al. (1971). The symbol  $\delta$  stands for boundary layer thickness.

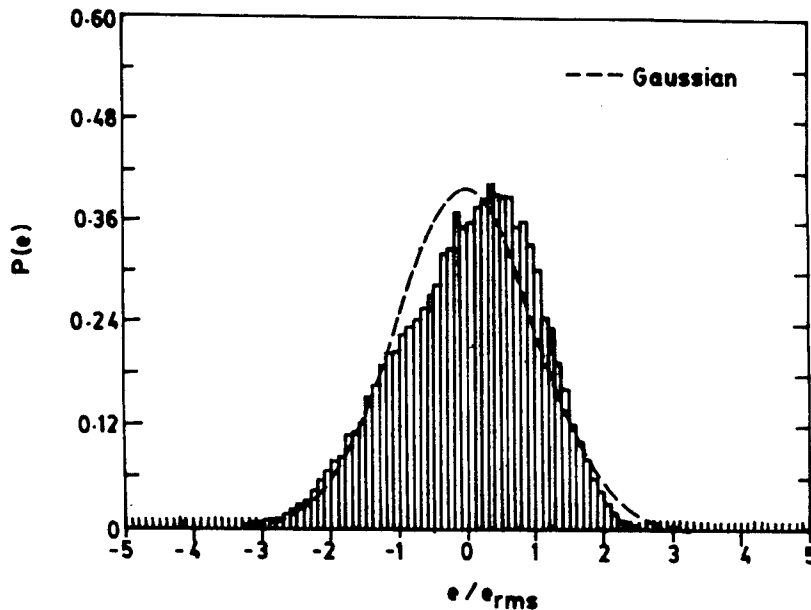
It was possible to compute the mean duration,  $\bar{T}_1$ , and the standard deviation  $\sigma_1$  of both the energetic and quiescent events separately. As shown in Table 1 the mean duration of energetic events was  $33.6 \text{ ms}$  compared to  $40.4 \text{ ms}$  for the quiescent events. Further, the quantity  $[e^2]_b / \overline{e^2}$  for the energetic events and quiescent events was computed to be 1.574 and 0.522, respectively. This means that the mean variance of energetic events is 1.574 times the long-time variance  $\overline{e^2}$ , and that of the quiescent events is 0.522 times the long-time variance.

The fraction of total time  $\gamma$  occupied by the energetic events was computed to be 0.454. Thus, the fraction of total time occupied by quiescent events was  $1 - 0.454 = 0.546$ . The fraction of one dimensional mean square turbulent burst energy was computed as  $\beta = \gamma [e^2]_b / \overline{e^2}$ . Similarly  $\beta$  could be computed for quiescent events using the corresponding values of  $\gamma$  and  $[e^2]_b / \overline{e^2}$  as shown in Table 1. Kim et al. (1971) measured  $\beta = 0.75$  and  $\gamma = 0.57$  for the bursting events in a turbulent boundary layer at  $y^+ \approx 15$ ,  $R\theta \approx 650$ . These values can be compared with  $\beta = 0.715$  and  $\gamma = 0.454$  for the energetic events in Table 1.

**Table 1: CHARACTERISTICS OF ENERGETIC AND QUIESCENT EVENTS**

	Energetic	Quiescent
$\bar{T}_1$	33.6 ms	40.4 ms
$\sigma_1$	20.3 ms	24 ms
$\bar{T}_1 U_\infty / \delta$	1.68	2.02
$\bar{T}_1 u_*^2 / \nu$	84	101
$[\overline{e^2}]_b / \overline{e^2}$	1.574	0.522
$\beta$	0.715	0.285
$\gamma$	0.454	0.546

The probability density functions for the total signal  $e(t)$ , the signal corresponding to energetic events  $e_b(t)$ , and the signal corresponding to quiescent flow  $e_q(t)$  are shown in Figure 3. The addition of 22,771 samples of Figure 3c and 27,379 of Figure 3c results in a total 50,150 samples, which is 26 samples short of a total of 50,176. These 36 samples appear to be lost due to numerical errors. Some very interesting observations can be made from these three figures considered together.



**Figure 3a: Probability Function of  $e(t)$ .**  
**Data Set = 50176 Samples**

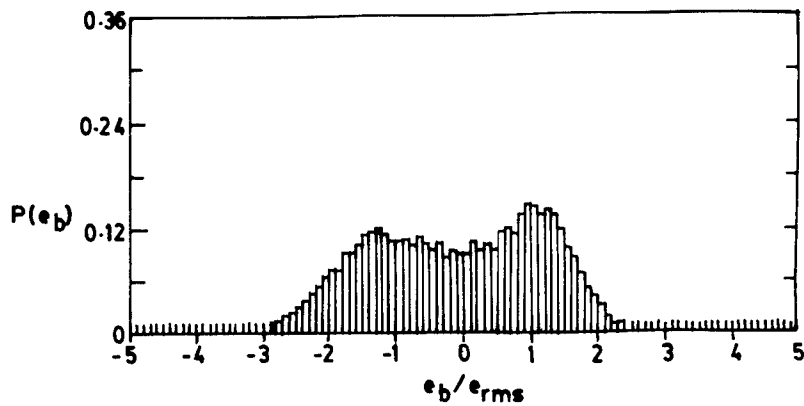


Figure 3b: Probability Function of  $e_b(t)$ .  
Data Set = 22771 Samples

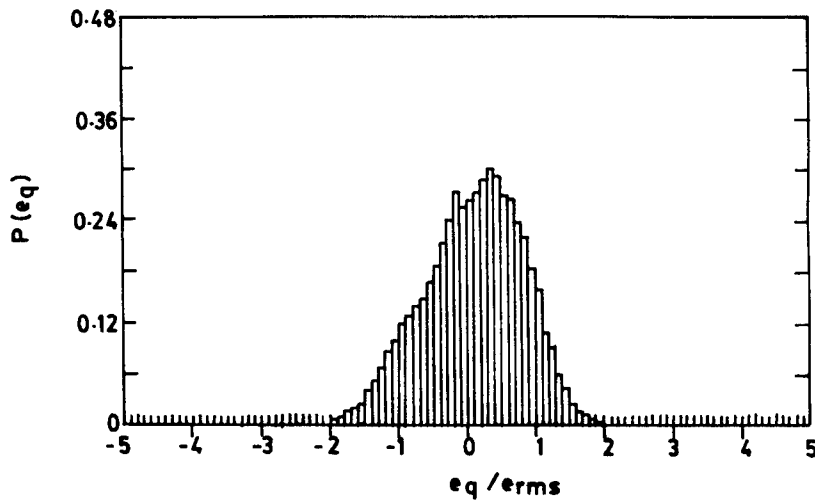


Figure 3c: Probability Function of  $e_q(t)$ .  
Data Set = 27379 Samples

First, the probability density function of quiescent flow  $e_q(t)$  modulates the probability density function of the total flow  $e(t)$  as the shapes of the curves  $P(e_q)$  and  $P(e)$  are similar. Second, the probability density function of energetic events  $P(e_b)$  shows a bi-modal distribution.

The important things to note are: i) the statistical features of the energetic events  $e_b(t)$  (Figure 3b) are completely masked in the statistical features of the total flow  $e(t)$  (Figure 3a); and ii) the probability density function  $P(e_q)$  of the quiescent events is not Gaussian and reflects the probability density function  $P(e)$  of the total turbulent fluctuations.

It was possible to further subdivide the energetic events into ejection-like events ( $u' > 0, v' > 0$ ) and sweep-like events ( $u' > 0, v' < 0$ ). For the present single hot-wire data set this subdivision was carried out by computing the short-time average value of each energetic event  $[\bar{e}_b]$ . Short-time average of energetic events with  $[\bar{e}_b] < 0$  was taken as indicating the ejection event, and  $[\bar{e}_b] > 0$  as the sweep event. Figure 4 shows this detection algorithm results, where the traces corresponding to  $e_{bj}(t)$  and  $e_{bs}(t)$  show the ejection and sweep events, respectively.

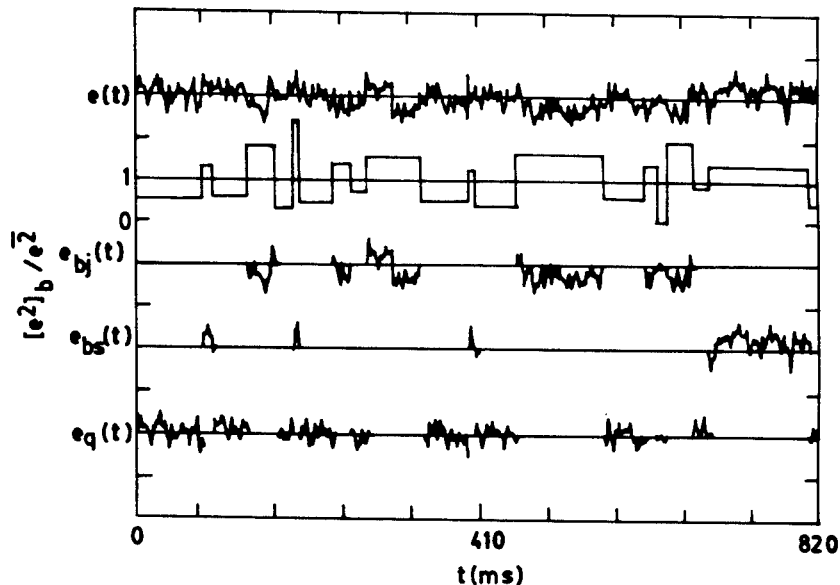


Figure 4: Detection of Ejection and Sweep Events;  
 $y^+ \approx 15$ . Data Set = 2048 Samples

The traces corresponding to  $e_{bj}(t)$  and  $e_{bs}(t)$  indicate that ejection and sweep-like events at a point in the flow do not always alternate. There are times when there appears at a point two ejection events in succession. This result of the detection algorithm is in agreement with flow visualization observations and also makes sense physically because the occurrence of energetic events is not local in space. There is a



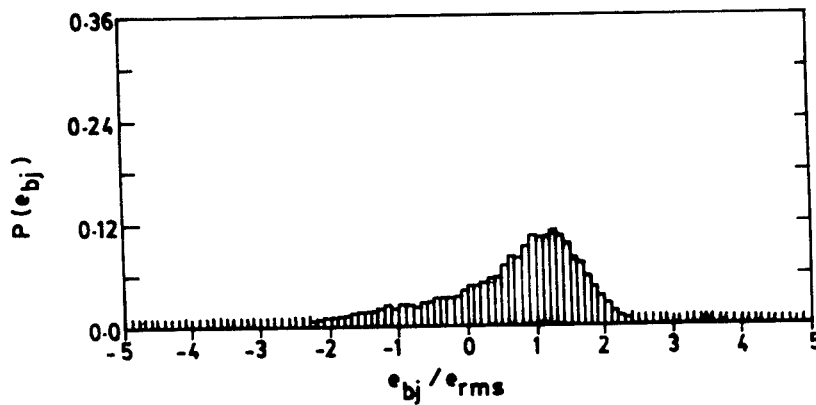
certain volume which each energetic event occupies as it occurs and moves downstream.

**Table 2: CHARACTERISTICS OF EJECTION AND SWEEP-LIKE ENERGETIC EVENTS**

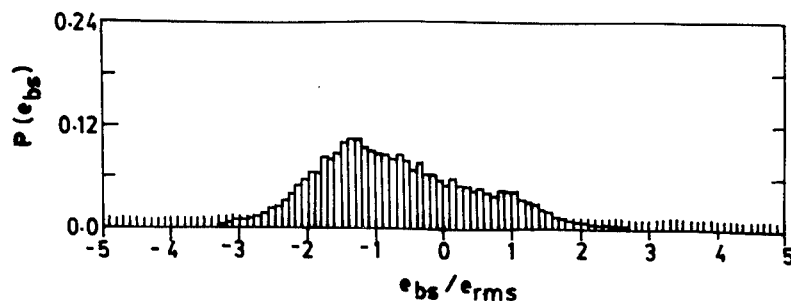
	Ejection	Sweep	Energetic (Table 1)
$\bar{T}_1$	32.8 ms	34 ms	33.6 ms
$\bar{T}_1 U_\infty / \delta$	1.64	1.70	1.68
$\bar{T}  u_* ^2 / \nu$	82	85	84
$ e^2 _b / e^2$	1.726	1.386	1.574
$\beta$	0.434	0.281	0.715
$\gamma$	0.251	0.203	0.454

Table 2 presents the statistical mean characteristics of ejection-like and sweep-like events. Out of a total of 271 energetic events for the present data set, there were 152 ejection-like events and 110 sweep-like events.

The probability density functions of the ejection-like events and sweep-like events are presented in Figure 5 which shows some remarkably distinct statistical properties of ejection and sweep events. While both  $P(e_{bj})$  and  $P(e_{bs})$  are highly skewed, the ejection events are positively skewed, and sweep events are negatively skewed.



**Figure 5a: Probability Function of  $e_{bj}(t)$ .  
Data Set = 12604 Samples.**



**Figure 5b: Probability Function of  $e_{bs}(t)$ .**  
**Data Set = 10167 Samples.**

Several other statistical characteristics of energetic and quiescent events as well as their spectral characteristics were also computed. Similar computations need to be carried out across the entire thickness of the turbulent boundary layer to establish the detection algorithm on firm ground.

### **3. Variable Interval Spectral Averaging (VISA) in a Backward Facing Step Flow.**

The HP 35660A dynamic signal analyzer digitizes a time record of 1024 samples for one channel to generate 512 points in the frequency domain by Fast Fourier Transform (FFT) and displays the first 401 points while discarding the rest. As many as 20 different frequency spans (digitization rates) ranging from 0.2 Hz (1 sample every 2 sec) to 102.4 KHz (262144 samples per second) are available. Furthermore, the number of time records (NTR) used for averaging of spectra can be chosen from 1 to 99999 time records and displayed on the monitor as well as stored on a floppy disk either sequentially with a constant number of time records or cumulatively with number of time records increasing in a cumulative manner. The contiguity is satisfactory for record lengths larger than the FFT processing time. A typical real time bandwidth for one channel mode is a frequency span of 800 Hz with averaging off and a frequency span of 3.2 KHz with fast averaging. All these characteristics are taken from the operating manual of HP 35660A dynamic signal analyzer.

The experimental conditions of test were a backward facing step of 25.4 mm height ( $h$ ) formed on the floor of a low speed wind tunnel of test section height of 304.8 mm and width of 405 mm. The free stream wind speed was 2 m/s giving  $Re_h = Uh/\nu \approx 2940$  with laminar separation and turbulent reattachment. A single hot-wire was located in the recirculation region at  $x/h = 5.5$  and  $y/h = 0.5$ . The signal from DISA 55A01 constant temperature anemometer was fed in parallel to a Kikusui

storage oscilloscope DSS 5020A and an HP 35660A Dynamic Signal Analyzer and plots were made on a HP Color Pro Plotter.

Since the objective was to detect low frequency oscillations, the selection of frequency span was the first step. By looking at the time traces and power spectra for  $NTR = 1$  and  $NTR = 250$  at frequency spans of 25 Hz, 200 Hz, 1.6 KHz and 3.2 KHz, it was decided that a frequency span of 200 Hz was most suitable for the present analysis. Figure 6 shows a typical time trace and power spectra for  $NTR = 1, 5$  and  $250$  for the 200 Hz frequency span.

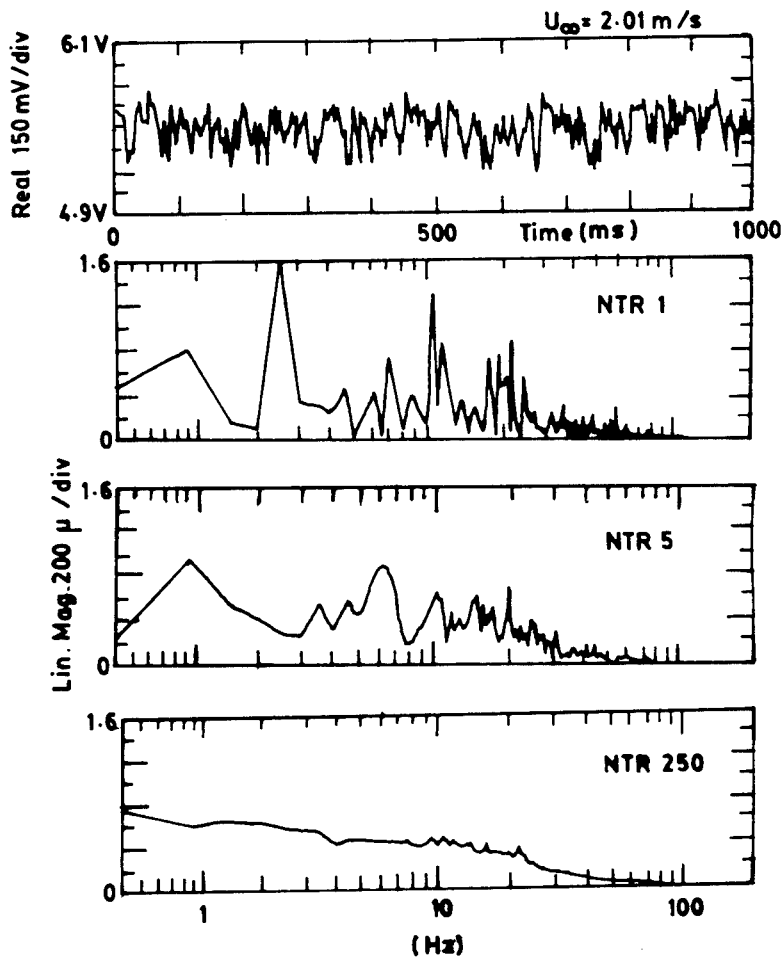


Figure 6: A Time Trace and Power Spectra for 200 Hz Frequency Span;  $NTR = 1, 5, 250$ .

Cumulative short-time average power spectra for  $NTR = 5, 10, 20$  and  $50$  are presented in Figure 7 to show the effect of averaging on the shape of the power spectral curve. As is to be expected, the increasing  $NTR$  smoothens the spectral curve without displaying conspicuous peaks due to averaging.

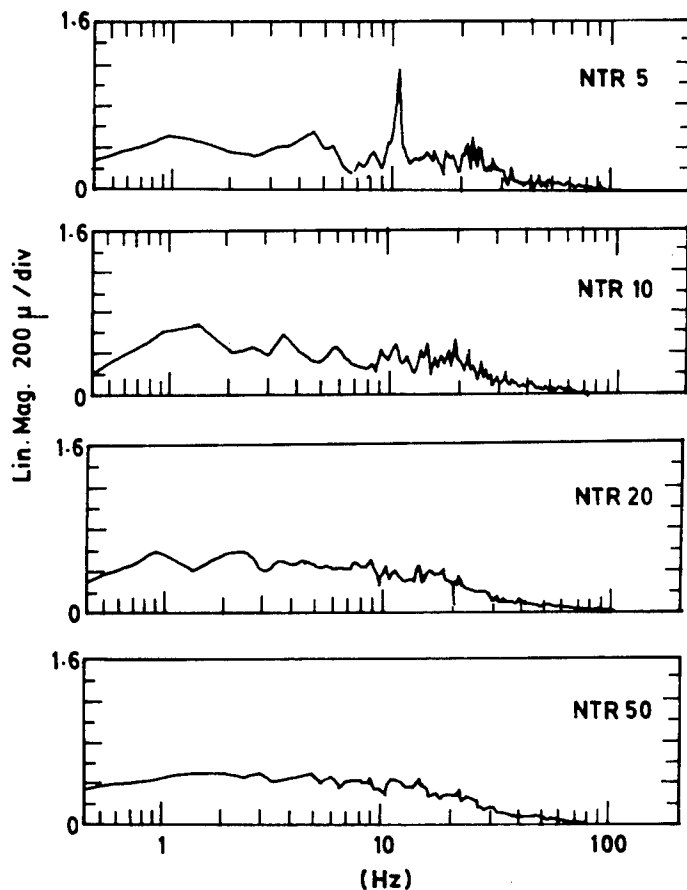


Figure 7: Cumulative VISA for  $NTR = 5, 10, 20$ , and  $50$ .  
Frequency Span =  $200$  Hz.

Figure 8 shows power spectra of four contiguous time records each of  $NTR = 5$ . It may be observed that low frequency peaks are clearly

discernible but there is no repeatability of peaks between any two spectra. Most of the dominant peaks are observed below 10 Hz.

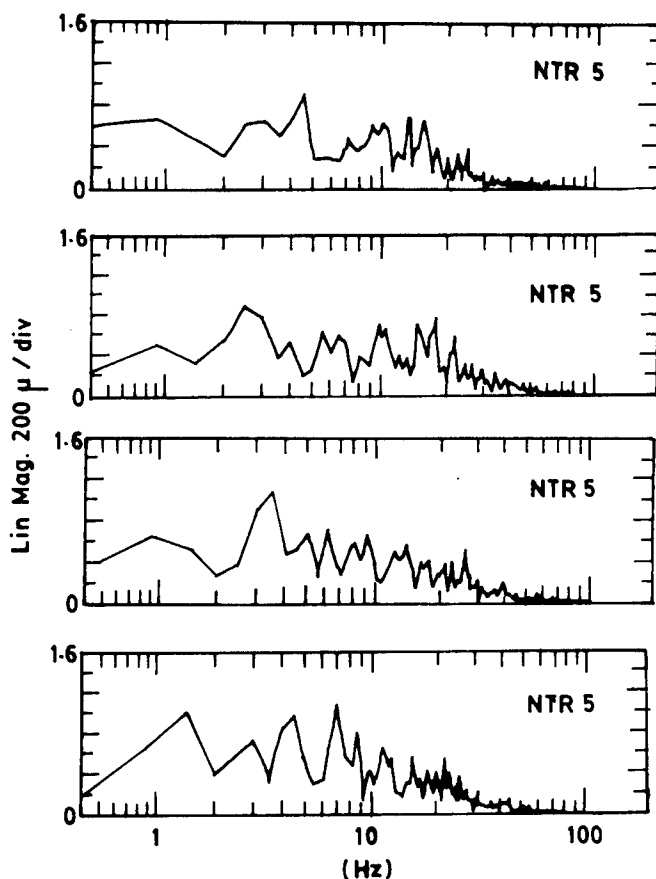
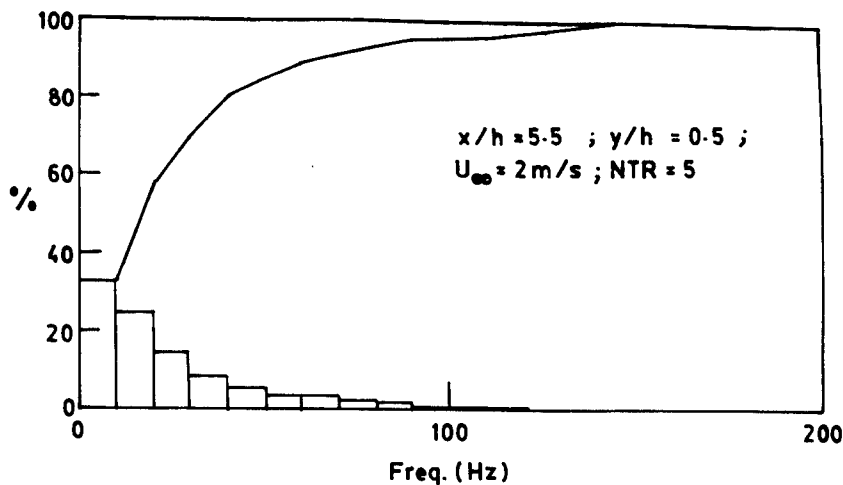


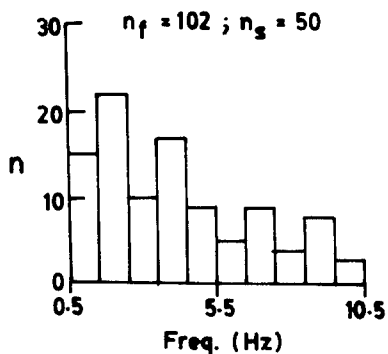
Figure 8: Typical DISA; NTR = 5. Frequency Span = 200 Hz.

Stepwise and cumulative energy curves averaged over 10 records of NTR = 5 in 10 Hz intervals are shown in Figure 9. This indicates that one third of the total energy is in the low frequency range of 0-10 Hz. The value of 10 Hz corresponds to  $St_h = 0.127$  and  $St_\theta = 0.006$ , where  $St_h = nh / U_\infty$ ,  $St_\theta = n\theta / U_\infty$ ,  $h = 2.5$  cm is the step height and  $n$  is the frequency in Hz.



**Figure 9: Stepwise and Cumulative Averaged Energy Distribution; NTR = 5. Frequency Span = 200 Hz.**

In the absence of repeatability of peaks, even in 0-10 Hz for the short-time averaged spectra ( $NTR = 5$ ), a histogram of peak frequencies in the 0-10 Hz range was made for 50 records of short-time averaged spectra ( $NTR = 5$ ). This histogram is shown in Figure 10. No single dominant frequency is apparent from this histogram. Instead, a group of frequencies is present more often than others. For the present case they are: frequencies of 2 Hz, 4 Hz, and 1 Hz, in the descending order of the histogram values.



**Figure 10: Frequency Histogram of Peaks; NTR=5. Frequency Span = 200 Hz.**

#### 4. The Structure of Turbulence

The experimental investigations reported in this paper are concerned with the structure of turbulence in turbulent shear flows. One of the turbulent shear flows considered is the turbulent boundary layer while the other is the turbulent separated flow. The objective has been to detect the coherent structures in the two cases: the burst by means of VITA algorithm, and the flapping oscillations by means of VISA algorithm.

Bursts are delineated by defining a cycle of fluctuation (about the conventional long-time mean) as a unit of short-time averaging. This makes the short-time averaging period a truly variable quantity, unlike the fixed short-time averaging period employed by most investigators with some amount of arbitrariness in selecting the fixed averaging time. A fluctuating cycle (about the conventional long-time average) can be regarded to represent an eddy, and a coherent structure can be regarded as made up of contiguous eddies. So far the definition of a turbulent eddy has been an enigma except for its identification with a scale of motion, and the belief that a turbulent eddy is not a wave (possibly a wave packet).

The detection of flapping oscillations in the separated turbulent flow on a backward facing step by means of short-time averaging applied in the frequency domain implies the identification of turbulent separated flow with the coherent structures. Most of the previous attempts to detect the flapping oscillations have been made by focusing attention in the free shear layer region of BFS flow. In the present case, the single hot-wire is located at  $y/h = 0.5$  in the recirculating region, and the flapping oscillations are regarded as due to overall coherent structure in the BFS cavity region.

#### 5. Conclusions

Two short-time averaging schemes applied to single hot-wire data are presented.

The first scheme employs a fluctuating cycle of flow oscillation about the mean as a unit of short-time averaging. With a threshold value chosen as the long time variance, the scheme is able to delineate the bursts and provide values for the burst frequency, intermittency, and other statistical characteristics of energetic events and quiescent flow at  $y^+ = 15$  in a turbulent boundary layer.

The second scheme considers the power spectra averaged over a few records for a frequency span of 200 Hz in the recirculation region of a backward facing step airflow at  $Re_h = 2940$ , and  $x/h = 5.5$ ,  $y/h = 0.5$  to detect the flapping oscillations. A group of frequencies of 2 Hz, 4 Hz and 1 Hz in the descending order of the histogram values is observed.

## References

- Ananda, B. K., *An Experimental Investigation of Structures of Flow Behind Backward Facing Step Using Spectra From Signal Analyzer* (Ph.D. Thesis, IIT Kanpur, 1993).
- Ananda, B. K., Gupta, A. K., & Oberai, M. M., *Proceedings International ASME Conference, Signals, Data, System*, **4**, Solid and Fluid Mechanics, India (1992) 49.
- Blackwelder, R. F., & Kaplan, R. E., *J. Fluid Mechanics*, **76** (1976) 89.
- Brown, G. L. & Roshko, A. J., *J. Fluid Mechanics*, **64** (1974) 775.
- Fluid Mechanics*, **30** (1967) 741.
- Farge, M., *Annual Review of Fluid Mechanics*, **24** (1992) 395.
- Gupta, A. K., Laufer, J., & Kaplan, R. E., *J. Fluid Mechanics*, **50** (1971) 493.
- Gupta, A. K., & Kaplan, R. E., *Proceedings Canadian Congress Applied Mechanics* (1987) B-126.
- Kim, H. T., Kline, S. J., & Reynolds, W. C., *J. Fluid Mechanics*, **50**, (1971), 133.
- Kline, S. J., Reynolds, W. C., Schraub, F. A., & Runstadler, P. W., *J. Fluid Mechanics*, **30** (1967) 741.
- Narahari Rao, K., Narasimha, R. & Badrinarayanan, M. A., *J. Fluid Mechanics*, **48** (1971) 339.



## SCALE-INVARIANT FORMULATION OF NONSTATIONARY SIGNALS

GEORGE TREVIÑO

Mechanical Engineering Department, Michigan Technological University  
Houghton, MI 49931 USA

E-mail: gtrevino@mtu.edu or 75244.2304@compuserve.com

A scale-invariant formulation of nonstationary signals is posed. The formulation engenders from the multiplicative group composition law of algebra rather than from the standard additive law. In particular, the explicit times at which signal values are correlated are converted into the principle coordinates of the formulation according to a nonlinear transformation which features the nonstationary character of the signal as a function of the geometric mean of the times rather than the standard arithmetic mean. Traditional time-lags are accordingly converted into time *ratios*. The advantages of this approach are discussed, particularly the related absence of correlation scales as well as the nonmetric aspects of the coordinate transformation defining the principal coordinates. Simple properties of second-order and third-order correlations are highlighted. A spectral version of the formulation is also developed, and it is shown that the *Mellin transform* is the transform best suited to this formulation.

### 1. Introduction

From the simplest perspective of engineering and/or physics, a nonstationary random signal is one in which at least one of its measurable statistics (e.g. *mode*, *median*, *mean*, *rms*, *skew*, etc.) is dependent upon the time-scale, say  $T$ , available for observation of the signal as well as on the initial time  $t$  of the observation. If the phenomenon under investigation is *nonlinear*, the stated  $t, T$ -dependence of one statistic typically guarantees a likewise dependence of another, and so on. In this respect nonstationary signals are different from stationary signals in that for a stationary signal, if  $T$  is "long enough," all of its measurable statistics are independent of both  $t$  and  $T$ . Occasionally, one may encounter a nonstationary signal whose measurable statistics are independent of either  $t$  or  $T$ , but such is the exception rather than the norm. There are some underlying features of nonstationary signals which are independent of both  $t$  and  $T$ , but these are not obvious, and exist largely in idealized cases. The present communication investigates one such case.

The theory of nonstationary random signals unfortunately still lacks firm and solid foundations. From a practical perspective, though, it is imprudent to procrastinate until such foundations are installed before applied analyses are attempted. Traditionally, the hesitation which the practical analyst expresses toward enjoining the problem(s) associated with nonstationary phenomena arises from the virtual absence of the appropriate necessary tools and/or methodologies for, and this next part

is critical, *describing and analyzing such phenomena in a mathematically simple way*. The recent advent of the theory of *wavelets* has enormously stimulated the interest of the signal processing community with respect to nonstationary analysis, and resulted in a virtual explosion in the number of related technical papers currently appearing in professional journals. In particular, the oft-stated contention that wavelets are more suited than are Fourier or other classical methods to the analysis and decomposition of nonstationary random processes has aroused the curiosity of more than one investigator and his (her) concomitant funding source. Wavelets have indeed on many instances successfully been applied to the analysis of nonstationary signals; but it appears, at least at this point in time, in a way which does not readily lend itself to the extraction from the results of such analyses those parameters that allow the analyst to unilaterally quantify, in a practical and meaningful way, the explicit nature of the attendant nonstationary behavior. Recall that from an applied science approach, the most fundamental intent of signal analysis is to determine from same a sufficient number of the defining statistics of the signal to effect some design-related decision. These statistics, keep in mind, are in most cases time-varying and scale-dependent. Quantifications of this nature, economic in the sense that they reduce the many to the few rather than increase the few to the many, as wavelet analysis seems to do, are the ones which in practical endeavors are typically held in highest regard. Wavelets and wavelet analysis, though, should eventually find their proper place in the literature as an advanced tool for signal analysis provided it can be established which wavelets are best suited to the various types of nonstationarities which manifest in everyday scenarios. Elsewhere in this volume (Andreas and Treviño, 1996) is presented a formulation consistent with this intent.

To somewhat alleviate the current state of affairs, then, the attention of this communication is directed to the study of the correlation structure of a certain fundamental class of nonstationary signals. One whose related characteristics are broad, but whose defining structure can nonetheless be simply modeled. The motivation is supplied by the belief that understanding a given phenomenon is always preceded by a simple and concise elucidation of same, coupled eventually with careful exploration, and exploitation, of its consequences. Although not suggested by the title, the intended class is the by-product of what could be termed "the natural pursuit of scaling," though perhaps a more appropriate designation is the "natural pursuit of re-scaling." The objective is to fully understand the analytical consequences of its particularly simple features, especially the imposed symmetry, and simultaneously evoke a concept of nonstationary behavior which is in some sense positive, with an accompanying degree of determinism.

## 2. The Nature of Nonstationarity

As discussed above, nonstationary random behavior is behavior some or all of whose defining statistics are explicitly time-dependent. "Time-dependent," again, means time-dependent in the sense of "instantaneous time" and in the sense of "time-scale.". Specifically, if the *histogram* of a nonstationary random signal is tabulated for the total number of data values occurring over a time interval defined by  $[t, t+T]$ , the defining statistics of the said histogram will vary with both  $t$  and  $T$ . It is assumed, by the way, that the measured signal is of some continuous ongoing physical phenomenon which essentially has no beginning and no end. Note that for a stationary process if  $T$  is "long enough," the shape of the histogram is independent of  $t$  and only the "number of occurrences" increases with increasing  $T$ . That is, the graph of number of occurrences vs signal values is only magnified along the ordinate by a factor which depends upon the increase in  $T$  only, and not on the specific value of the signal. Actually, such a dependence with increasing  $T$  in the shape of the histogram is perhaps more definitive of an *ergodic* process than it is of a stationary process, but in everyday practice *ergodicity* and *stationarity* are invoked interchangeably. At once, then, the "two-dimensional" nature of nonstationary behavior becomes obvious. All physically meaningful statistical characteristics of the behavior depend on both instantaneous time and total time of measurement/observation. A stationary process, on the other hand, is zero dimensional in that if the measurement time is long enough, all meaningful statistical characteristics of the behavior are independent of both time and scale. The histogram, recall, is a depiction of the number of relative occurrences of the the signal values over one finite event of the process.

Nonstationary behavior does not readily reveal what could be designated as "universal structure," and is accordingly not what would normally be considered an idealized type of behavior. In the classical formulation of the multi-time correlations peculiar to nonstationary analysis, one of the more elementary decisions an analyst must make in order to quantify the temporal evolution(s) of same is the choice of appropriate time scales with which to nondimensionalize the various time lags that typically appear in the correlations. Time scales for the temporal evolutions of the rms, skew, kurtosis, etc. are typically defined by the process itself, and in this sense are natural to the phenomenon. The analyst is (usually) not at liberty to modify these. Multi-time correlations, recall, are integral inner-products which the analyst uses to establish a relationship between the behavior of a random signal at one time instant with that at another. The proper choice of time scale for time-lag nondimensionalization, unfortunately, has never been fully addressed in the literature, and consequently a variety of differing time scales have historically been introduced to satisfy particular conditions that arise in specific analyses. One such scale is defined as

$$\eta(t, \tau) = - \left[ \frac{\partial^2 R(t, \tau)}{\partial \tau^2} \right]^{-0.5}, \quad (1)$$

while a second is

$$\Lambda(t) = \int_0^{\infty} R(t, \tau) d\tau. \quad (2)$$

An interesting feature of the time scale defined by Equ. (1) is its  $\tau$ -dependence, which essentially allows the "larger" scale behavior of the considered signal to be scaled in a manner different from the "smaller" scale behavior ( $\sim$  "scale-dependent" scaling, sometimes denoted "generalized" self-similarity). In both of these definitions  $R(\cdot)$  is the two-time correlation of the given random signal normalized by the square of its rms. Scales peculiar to third, or even higher, -order, correlations can also be defined, as can the appropriate normalization factors. In the analysis of "turbulent" fluid phenomena, other time dependent scales such as the *Kolmogorov* scale and the *dissipation* scale are not uncommon. The dissipation scale of turbulence is analogous to the scale defined in Equ.(1) evaluated, though, at  $\tau = 0$ , while the Kolmogorov scale is that scale which makes the turbulence *Re* approximately equal to unity.

The particular choice of time scale is important because it is felt by many investigators that when random signals are "properly" scaled, they reveal features which are independent of how the signal is physically measured. Features which are independent of how a signal is measured, one might suspect, suggest a universality of sorts, which in turn suggests properties that are manifestations strictly of the mechanisms of the random behavior itself and not of any physical mechanism(s) which create(s) the behavior. A sort of "self-governance", if you will, much like the thermal efficiency of a Carnot cycle in thermodynamics which, as is well known, is independent of the working fluid. The isolation of such autonomy in random behavior is potentially the beginning of the formulation of any law, or laws, that define or describe the related underlying physics.

The need for a time scale in the classical formulation of nonstationary random signals arises because it is both natural and convenient to model the temporal evolution of, say, the standard two-time correlation,  $Q(t_1, t_2) = \langle x(t_1)x(t_2) \rangle$ , of such a signal in terms of a time-dependent time scale as  $C(t, \tau) \approx \sigma^2(t)R(\tau/L)$  where,  $t = (t_1 + t_2)/2$ ,  $\tau = (t_2 - t_1)$ , and  $L$  is the chosen time-scale which is in some sense characteristic of the second-order correlation structure of the signal, and almost invariably varies with  $t$ . Consistent with accepted notation, angular brackets,  $\langle \rangle$ , denote ensemble averaging and  $\sigma(t)$  denotes the time-dependent rms of the signal. This approximation is equivalent to writing the measured

random signal itself as the product  $\mathbf{x}(t) = \sigma(t)\zeta(t)$ , and ultimately in fully nondimensionalized form as  $\mathbf{x}(t) = \sigma(t)\zeta(t/L)$ . The premier feature of this latter expression is the roles of both  $\sigma$  and  $L$  as time dependent scaling parameters. The  $t$ -dependence of  $L$ , again, is required because classical nonstationarity in its most complete form demands that the correlation structure ( $\sim$  "degree of randomness", frequency content, etc.) of the normalized random process  $\zeta(t)$  also vary with time. In other words, there is no guarantee that for purposes of second-order correlation analyses, the complete averaged time-behavior of  $\mathbf{x}(t)$  is captured by  $\sigma(t)$  only. For third-order correlations the signal is instead decomposed as  $\mathbf{x}(t) = \nu(t)\xi(t/M)$  where  $\nu$  is the skew of the process and  $M$  is a time-dependent time scale in some sense characteristic of  $\xi(\cdot)$ .

### 3. Examples of Nonstationary Behavior

Perhaps the simplest example of nonstationary behavior is that of observing and measuring over a short time-interval one member of a random process, say  $\mathbf{s}(t)$ , which is known to be stationary. Denoting the measurement time-interval as  $T$ , the recorded signal is thus  $\mathbf{x}(t) = [U(t) - U(t-T)]\mathbf{s}(t)$ , where  $U(t)$  is the *unit-step* function. The two-time correlation function of the recorded signal is accordingly

$$\langle \mathbf{x}(t_1)\mathbf{x}(t_2) \rangle = \langle \mathbf{x}(t - \tau/2)\mathbf{x}(t + \tau/2) \rangle = C_{\mathbf{x}}(t, \tau) = [U(t - \tau/2) - U(t - T + \tau/2)]C_{\mathbf{s}}(\tau), \quad (3)$$

where in a practical scenario the correlation of particular interest is invariably  $C_{\mathbf{s}}(\tau) = \langle \mathbf{s}(t-\tau/2)\mathbf{s}(t+\tau/2) \rangle$  rather than the computed  $C_{\mathbf{x}}(t, \tau)$ . Recall that in practice,  $\langle \mathbf{x}(t-\tau/2)\mathbf{x}(t+\tau/2) \rangle$  is commonly approximated as

$$q(\tau, T) = T^{-1} \int_0^T \mathbf{x}(t - \tau/2)\mathbf{x}(t + \tau/2) dt. \quad (4)$$

The situation at hand, then, is clearly a special case of the general problem,  $\mathbf{x}(t) = w(t)\mathbf{s}(t)$  where  $w(t)$  is a known *window function* typically of some windowing scale,  $T$ . A straightforward computation for the general case produces  $C_{\mathbf{x}}(t, \tau) = w(t-\tau/2)w(t+\tau/2)C_{\mathbf{s}}(\tau)$ , indicating that the two-time correlation of the recorded signal is unavoidably a  $t$ - and  $\tau$ -dependent modulation of  $C_{\mathbf{s}}(\tau)$ . For values of  $\tau$  small in comparison to  $T$ , say roughly  $(\tau/T) < 0.1$ , the difference between  $C_{\mathbf{x}}(t, \tau)$  and  $C_{\mathbf{s}}(\tau)$  is minimal, but for correspondingly large values of  $\tau$  the difference can be substantial. The exact net effect, though, is best analyzed by means of frequency analysis. A very fundamental "hardware-oriented" analysis of how the frequency content of a short-lived signal is distorted by a transient recorder is presented by Leighton (1988).

Another example of a nonstationary process is a phenomenon known as  $1/f$  noise. The accepted definition of  $1/f$  noise is noise whose power spectrum  $\phi(f)$  is of the form  $\phi(f) \approx (\text{constant}) / f^{-\gamma}$ , where  $0 < \gamma < 2$  and  $f$  denotes circular frequency; invariably, though,  $\gamma = 1$ . Because the measured signal, say  $X(t)$ , is nonstationary, the term "power spectrum" is invoked here in a rather "loose" sense, and what is really intended by this designation is instead the absolute value of the *periodogram*,

$$Z(\omega, T) = \frac{1}{\sqrt{T}} \int_0^T X(t) \exp(-i\omega t) dt, \quad (5)$$

where  $\omega$  is the angular frequency, equal to  $2\pi f$ . The "power spectrum" of the signal is then found according to  $\phi(f) \sim \phi(\omega, T) = Z(\omega, T)Z^*(\omega, T)$ .

Lastly, an indeed very interesting example of nonstationary behavior is the case of turbulence generated by a grid in a wind or water tunnel, or some other such mechanism, which then decays as it is advected by the mean flow away (downstream) from turbulence generating source. Such turbulence is usually denoted as *nonhomogeneous* steady-state turbulence because its statistical characteristics vary with spatial position and not with time; but a  $t$ -dependence can be introduced into the behavior by invoking the "frozen" turbulence assumption, sometimes known as Taylor's hypothesis, allowing downstream distance from the source to be approximated as  $\bar{U}t$ , where  $\bar{U}$  denotes the mean flow. In this way, spatial variations in the stochastic averages of the turbulence can be instead modeled in terms of analogous time-dependencies.

#### 4. Invariance in Nonstationary Phenomena

The concepts of invariance (Salmon, 1885; Grace and Young, 1903; Gurevich, 1964) have quietly been applied to the analysis of random processes since the late thirties. Publishing in the 1904-05 issue of the *Hilbert Journal*, C. J. Keyser describes invariance as "changelessness in the midst of change, abiding things in a world of flux, configurations that remain the same despite the swirl and stress of countless hosts of curious transformations." The importance of invariance has already proved itself in such diverse fields as algebra, geometry, theoretical physics, and psychology, and its use as a bonafide tool of thought is exemplified by the classical notion of conservation of energy. In fluid mechanics this law is accompanied by conservation of mass, conservation of linear momentum, and conservation of angular momentum. Scientists, in their unending quest for knowledge, typically search for invariant properties whether they know it or not, such being simply a manifestation of the natural aspiration for generality. In practically all sciences the simplest principles that simultaneously have the broadest application are the ones which are most valued. Invariance

is most subtle in that it asserts that the substantive laws of a given field should be expressible in a manner which is strictly independent of the particular scales chosen to measure the related field variables. The *Hibbert Journal*, by the way, is a "quarterly review of religion, philosophy, and theology;" it was discontinued in 1968.

The existence of invariant structure in random behavior has been suspected for some time. In fact, invariance underlies and is intimately connected to the very definition of *stationarity* itself. To be sure, a stationary random signal is defined as one whose *probability density function* (pdf) is independent of arbitrary translations in time, and therefore one whose statistical structure is independent of the particular choice of reference origin. This well-known translation invariant formulation of random signal analysis has its roots in the principle of theory construction that since the laws of mechanics are invariant under uniform translations (and rotations) of the underlying coordinate system, the phenomenon under investigation should likewise reveal similar features. A transformation of the type,  $t' = t + \varepsilon$  where  $\varepsilon$  is a constant, is perhaps the most fundamental of those transformations which reflect this property, but a transformation of the type  $t' = At + \varepsilon$  where  $A$  is a constant also reflects the translation invariant (but scale-dependent) property. Translation invariance implies that the very idea of "origin" is an illusion, acknowledging only a "before" and an "after," but not when this distinction occurs. Nonstationary processes, however, are not invariant with respect to these types of transformations, and the time-dependencies of  $\sigma$  and  $L$  indicated above clearly establish that as time evolves,  $C(t + \Delta t, \tau) \neq C(t, \tau)$ . This accordingly suggests that nonstationary processes are perhaps invariant under a more general type of transformation.

In order to introduce, then, the concept(s) of invariance into the analysis of nonstationary signals, it is necessary to invoke a more general type of invariance; the intended invariance is *scale invariance*. The ensuing formulation, in a somewhat less restrictive form, is related to the formulation of Gray and Zhang (1988).

## 5. Scale-invariant Formulation

Consider a nonstationary random signal, denoted by  $x(t)$ , whose  $n$ -th order probability density function (pdf) is  $p\{x(t_1), x(t_2), x(t_3), \dots, x(t_n)\}$ . For purposes of first- and second-order moment analysis only, suppose that  $p\{x(t)\} = p\{x(At)\}$ , and  $p\{x(t_1), x(t_2)\} = p\{x(At_1), x(At_2)\}$ , where  $A$  is potentially a time-dependent scale change. Since for this type of transformation there is no invariant in the instantaneous time  $t$ , it follows that the *mean* of such a process, in order to satisfy the required invariance, must be a constant, which for practical purposes can be taken to be zero. When  $A$  is a constant, though, the two-time correlation is a function of  $\alpha = t_2/t_1$  only, since this is the only invariant in two times under this type of scale

change transformation. Thus for such a "scale-invariant process,"  $\langle x(t) \rangle = 0$  and  $\langle x(t_1)x(t_2) \rangle = Q(\alpha)$ . This is the result of Gray and Zhang referenced above, but they called their formulation *M-stationary*. Actually, this simple type of invariance is perhaps too restrictive for many types of nonstationary signals encountered in practice, and what is required in order to relax the invariance is to first determine the square of the signal rms according to  $\langle x^2(t) \rangle = \sigma^2(t)$  and then re-write  $x(t)$  in normalized form as  $\sigma(t)\zeta(t)$ , where  $\zeta(t)$  is a random signal of unit rms, and pdf  $\varpi\{\zeta(t_1), \zeta(t_2)\}$ . Imposing now the indicated scale-change invariance on the pdf of  $\zeta(t)$  forces this signal, consistent with  $x(t)$ , to also have a zero mean, and a second-order correlation, call it  $R$ , which is still a function of  $\alpha$  only. In this approach, the time-dependence of the signal rms, unlike the more restrictive invariance, is not lost, a rather key ingredient to nonstationary analysis, and furthermore the need to introduce a time-scale into the two-time correlation function is avoided. Since  $R(\alpha)$  is independent of scale, it is *fractal* in structure.

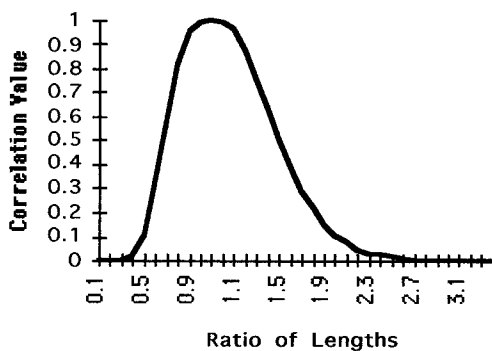
This invariance can be relaxed further yet, if it is possible to invoke invariance under a time-dependent scale-change such that the related scale-change for second-order moments is "approximately" the same for both  $t_1$  and  $t_2$ . For example, suppose that  $\varpi\{\zeta(t)\} = \varpi\{\tilde{\zeta}(t)A(t)\}$ , where  $A(t)$ , as indicated, is a time-dependent scale-change. Under this type of invariance, the mean of  $\zeta(t)$  is still a constant ( $= 0$ ), but if the scale-change is such that  $\varpi\{\tilde{\zeta}(t_1)A(t_1), \tilde{\zeta}(t_2)A(t_2)\} \approx \varpi\{\tilde{\zeta}(t)A(t_1), \tilde{\zeta}(t)A(t_2)\}$ , where  $t$  is some "average" value between  $t_1$  and  $t_2$ , say  $t = (t_1 t_2)^{1/2}$ , or maybe even  $t = (t_1 + t_2)/2$ , then the second-order correlation, again, **is still a function of  $t_2/t_1$  only**. The important feature for this invariance to manifest is that there exist a time-dependent scale which is "common" to both  $t_1$  and  $t_2$ . The knowledge that certain properties of a particular physical phenomenon actually remain preserved during a considered evolution, in any sense, can be of great help not only in simplifying any governing equations, but also in eventually leading to their solution. At the very least it can lead to new and different insights.

An immediate advantage of this formulation is that for purposes of second-order moment analyses, all nonstationary behavior in  $x(t)$  is carried, as it should be, by the time-dependent rms of the signal and not by the correlation structure. Recall that the absolute value of the correlation between  $\tilde{\zeta}(t_1)$  and  $\tilde{\zeta}(t_2)$  is a number greater than or equal to zero and less than or equal to unity which is a measure of that fraction of  $\tilde{\zeta}(t_1)$  that "follows"  $\tilde{\zeta}(t_2)$  (or vice versa). Historically, correlations have been used by statistical analysts to determine a number of statistics of random signals which are different and distinct from the classical *mode*, *median*, and *mean*. A possible disadvantage of this formulation is that the imposition of scale-invariance has come at the expense of translation invariance. Translation invariance (stationarity), though, has never had much practical significance, since most random behavior



encountered in nature is not of the translation invariant variety and moreso, in a realistic scenario, the location of the origin of coordinates (time in this case) during any physical measurement of random behavior is invariably held fixed, and therefore "privileged," by the analyst. Scale invariance acknowledges only "larger than" and "smaller than," but not where one ends and the other begins. While translation invariance focuses on motion through time and has no preferred origin, scale-invariance focuses on motion through scales and has no preferred scale. Random behavior typically encountered in nature is exactly this, since at any time instant it can be safely assumed that a multitude of scales manifest in the behavior, with none of these scales ever ascending to a "preferred" status.

FIGURE 1: Scale-Invariant Correlation



Simple properties of the formulated invariance are  $R(\alpha) = R(\alpha^{-1})$ , and  $R(1) = 1$ . These are respectively analogous to  $R(\tau) = R(-\tau)$  and  $R(0) = 1$  of the classical formulation. These two properties suggest that  $R(\alpha)$  in the scale-invariant formulation has a geometric shape similar to that depicted in Figure 1. For third-order correlations, the corresponding symmetry properties are  $S(\alpha_1, \alpha_2) = S(\alpha_2, \alpha_1) = S(\alpha_1^{-1}, \alpha_2/\alpha_1)$  where  $\alpha_1 = t_2/t_1$ ,  $\alpha_2 = t_3/t_1$  and  $Q(t_1, t_2, t_3) = \sqrt{3}(t) S(\alpha_1, \alpha_2)$ .

Here  $t = (t_1 t_2 t_3)^{1/3}$  and note that  $S(1, 1)$  has no a priori defined value, since third-order correlations are zero for gaussian signals and non-zero for any signal which has a pdf non-symmetric about its *mean*. These rather peculiar symmetry properties impose upon  $S(\alpha_1, \alpha_2)$  a distinctly non-arbitrary algebraic form. Specifically,  $S(\alpha_1, \alpha_2)$  must be expressible in terms of the forms,  $\alpha_1 \alpha_2 / (\alpha_1 + 1 + \alpha_2)^3$ ,  $\alpha_1 \alpha_2 / (\alpha_1 + \alpha_1 \alpha_2 + \alpha_2)^{3/2}$ , etc., since these forms (and others like these) automatically satisfy the required argument symmetry.

## 6. Generic Example

Of special interest to the practitioner, though, is "what types of nonstationary behavior encountered in 'real world' scenarios can ideally be characterized in this scale-invariant fashion?" Consider again, then, the random process  $x(t)$  discussed earlier which was later expressed as  $x(t)$

$\approx \sigma(t)\tilde{q}(t/L)$ . In order for a random signal to be reducible to this form it is necessary that its correlation function(s) adhere to a somewhat specialized algebraic structure. For example, normalization of the two time correlation with respect to the square of the rms is natural and deducible from the *Schwarz Inequality*. This produces in the "standard" notation the normalized correlation function  $R(t_1, t_2)$ , which in said notation re-writes as  $R(t, \tau)$  where  $t$  and  $\tau$  have been previously defined. In applied work it is not uncommon to let the  $t$ -dependence of  $R(t, \tau)$  be carried by a single time scale, say  $L(t)$ , or possibly even a  $t$ - and  $\tau$ -dependent time scale, say  $L(t, \tau)$ , such that  $R(t, \tau) \Rightarrow R\{L(t), \tau\}$  or  $R(t, \tau) \Rightarrow R\{L(t, \tau), \tau\}$ . Whichever mode is invoked, though, the two coordinates of interest, viz.  $L$  and  $\tau$ , can be converted into the two *principle* coordinates,  $\phi = (L\tau)^{0.5}$  and  $\theta = (\tau/L)$ , such that  $\tau = \phi/\theta$ ,  $L = \phi/\sqrt{\theta}$  and  $R\{L, \tau\} \Rightarrow R(\phi, \theta)$ . If the condition of scale invariance is now imposed on  $R(\phi, \theta)$ , it follows that it cannot be a function of  $\phi$ , since any (pure scaling) transformation which carries  $\tau \Rightarrow A\tau$  also carries  $L \Rightarrow AL$ , leaving  $\theta$  unaffected while taking  $\phi \Rightarrow A\phi$ . Under this constraint, the originally measured random signal can now be written as  $x(t) \approx \sigma(t)\tilde{q}(t/L)$  with the " $\approx$ " sign required since scale invariance has been artificially imposed on  $\tilde{q}(t)$ , and is ideally not an intrinsic feature of any normalized random signal. Note that  $x(t)$  is itself neither scale nor translation invariant since the time-dependent rms of  $x(t)$  is scale and origin dependent. Translation invariance, though, can be imposed in an "approximate" sense on the scale invariant  $\tilde{q}(\theta)$  such that the correlation between  $\tilde{q}(\theta_1)$  and  $\tilde{q}(\theta_2)$  is a function of  $\theta_{2c} - \theta_{1c}$

only, where  $\theta_{2c} = \frac{t_2}{L_c} \neq \frac{t_2}{L_2}$  and  $\theta_{1c} = \frac{t_1}{L_c} \neq \frac{t_1}{L_1}$ . This is possible only if

$$\frac{t_2}{L_2} - \frac{t_1}{L_1} \approx \frac{t_2}{L_c} - \frac{t_1}{L_c} = \frac{\tau}{L_c}, \text{ i.e. if a scale "common" to both } t_1 \text{ and } t_2 \text{ can be}$$

found such that the difference between the scale invariant coordinates scaled with different scales is approximately equal to the difference when the coordinates are scaled with the *same* scale ( $\sim$  "scaled stationarity"). Typically this "common scale" is defined either as  $(L_1 + L_2)/2$  or as  $(L_1 L_2)^{0.5}$ . In this way, "global" information, through scale and (approximate) translation invariance, and "local" knowledge, through scale and origin dependence become equal partners in a viable representation of random behavior, with neither attempting to deny the other by unduly extending its own boundaries. For such potential formulations, the correlation between  $\tilde{q}(\theta_1)$  and  $\tilde{q}(\theta_2)$  can also be written as  $R(\alpha) = R(\theta_2/\theta_1)$  as well as  $R(\theta_2 - \theta_1)$

## 7. Spectral Formulation

Any formulation of the theory of random signals would be incomplete without a spectral version. Spectral analysis, recall, resolves the signal

into components which make additive contributions to the energy in the signal and have clearly recognizable physical meaning. While the demand for additive contributions to the energy suggests that the signal be represented in terms of a set of orthogonal functions, it will be demonstrated momentarily that the obvious choice in this formulation is not the well-known trigonometric functions. In trigonometric analysis, it is the wavelength ( $\sim (\text{frequency})^{-1}$ ) of a given trigonometric function which explicitly identifies it and indeed isolates it from other members of the set. Ordinary Fourier methods correspond, therefore, to a resolution of the signal into components of different linear size. Additionally, trigonometric functions neither increase nor decrease their amplitudes at infinity, which is clearly not the case in nonstationary behavior. In the approach formulated herein, it is not "size" but rather *ratio* of sizes which defines the correlation structure of the signal.

The concept of a spectrum for nonstationary processes **is not a new concept**. It is in fact discussed in several advanced texts on the subject of random processes (see, for example, Bendat and Piersol (1971) and Papoulis (1965)). Recall that a "spectrum," in its most classical sense, is a fully time-independent function that describes the distribution of signal energy over frequency space. In this respect it is equivalent to the corresponding concept for the stationary case. Unfortunately, though, it is an idea whose acceptance, to date, is not widespread. One reason for this is due to the difficulty in finding a suitable set of orthogonal eigenfunctions which can be utilized to formulate a spectral decomposition for the general case of nonstationary behavior. Nonstationarity, by its very nature, implies phase consistency, phase coherence, etc., between pairs of Fourier frequencies, thus, as stated earlier, automatically excluding the use of standard complex exponentials as the orthogonal eigenfunction set. Another is that the essence of a spectrum for nonstationary phenomena is not completely clear, and on the surface appears to be somewhat paradoxical. If viewed properly, though, this paradox can be made to yield useful consequences, since there is virtually no paradox without some utility. Research on this topic has been done and indeed reported in the literature (Page, 1952; Lampard, 1954; Turner, 1954; Levin, 1964; Priestley, 1965 and 1967; Loynes, 1968; Nagabhushanam and Bhagavan, 1968; Bhagavan, 1974). Nonetheless, the concept of a spectrum for nonstationary process is still not an idea whose formulation is widely accepted nor whose functional structure has been completely investigated.

The concept is realized by *formally* extending, in a most natural sort of way, the fundamental relationship between the correlation and spectrum functions for the stationary case to obtain, for the nonstationary case, the expression,

$$G(t_1, t_2) = (2\pi)^{-2} \int \exp\{-i(\omega_1 t_1 - \omega_2 t_2)\} \Phi(\omega_1, \omega_2) d\omega_1 d\omega_2. \quad (6)$$

Here,  $\Phi(\omega_1, \omega_2)$  is the Fourier frequency spectrum for nonstationary processes, and is historically designated a *generalized frequency spectrum* or *generalized power spectrum*. The word "formally" is emphasized above since the time-dependent nature of second-order correlations for any nonstationary process requires that Equ. (6) cannot be an equality in general, i.e. complex exponentials, by their very nature, cannot be the natural eigenfunctions for any (save the most highly specialized) time-dependent dynamical system. Analysis of a simple spring-mass system with a time-dependent spring (or mass) immediately establishes this result.

In order to establish, though, through purely mathematical arguments, the existence of a spectrum function for nonstationary signals in general it is only necessary to adapt the results of a theorem peculiar to the spectral analysis of time series. The essence of the theorem is that any random process,  $\mathbf{z}(t)$ , even a nonstationary one, with a continuous correlation function has an orthogonal expansion with respect to some appropriate set of eigenfunctions  $\gamma_n(t)$ , i.e. for any  $\mathbf{z}(t)$ , it is always possible to write

$$\mathbf{z}(t) = \lim_{N \rightarrow \infty} \sum_{n=1}^N \alpha_n \gamma_n(t) \quad (7)$$

where the  $\alpha_n$  are random coefficients, *l.l.m.* denotes *limit in the mean*, and it can be shown that  $\gamma_n(t)$  and the  $\alpha_n$  are determined by solving the integral equation,

$$\int Q(t, s) \gamma(s) ds = \delta^2 \gamma(t), \quad (8)$$

where  $\delta$  is a constant. The condition for which there exists an infinity of solutions to Equ. (8) is a premise of the theory of integral equations, and will not be addressed here. The reader who prefers more discussion of this theory is referred to Chapter III of Courant and Hilbert (1954).

The spectral version of the theory posed here can be effected by introducing the variable  $\rho = \ln t$ ; or alternatively  $t = \exp \rho$ . This variable nonlinearly transforms the normalized random signal  $\zeta(t)$  into  $\varphi(\rho)$  according to  $\zeta(t) = \zeta(\exp \rho) \leftrightarrow \varphi(\rho)$ . In view of this relation, the correlation  $R(\alpha)$  rewrites as  $R(\exp \rho_2 / \exp \rho_1) \leftrightarrow D(\rho_2 - \rho_1)$ , where  $D(\rho_2 - \rho_1)$ , because it is a function of the difference,  $\rho_2 - \rho_1$ , is the correlation of the stationary random function  $\varphi(\rho)$ . That is, the effected transformation from  $t \leftrightarrow \rho$ , together with the nonlinear transformation from the physical coordinates  $(t_1, t_2)$  to the principal coordinates  $t = (t_1 t_2)^{1/2}$  and  $\alpha = t_2 / t_1$ , has converted a nonstationary signal whose classical normalized correlation function evolves in time in a manner which can be modeled by the introduction of a single time-dependent time scale, into a signal

whose normalized second-order correlation function is strictly stationary. Therefore,

$$\phi(\beta) = \int R(\alpha) \alpha^{-\beta-1} d\alpha, \quad 0 \leq \alpha \leq \infty \quad (9)$$

where  $\beta$  is a wave number and  $\phi(\beta)$  the Mellin transform of  $R(\alpha)$ . Baudelaire (1974) has shown that the Mellin transform plays a role in the analysis of scale-invariant systems similar to the role the Fourier transform plays in the analysis of shift-invariant systems, viz. it is a wave number spectrum such that the total kinetic energy per unit mass is found by summing  $\phi(\beta)$  over all component wave numbers. This utility manifests because of the property that the magnitude of the Mellin transform is insensitive to scaling of the independent variable. The inverse transform follows from

$$\begin{aligned} D(\rho_2 - \rho_1) &= (2\pi)^{-1} \int \phi(\beta) \exp\{i\beta(\rho_2 - \rho_1)\} d\beta \\ &= (2\pi)^{-1} \int \phi(\beta) \alpha^{i\beta} d\beta = R(\alpha). \end{aligned} \quad (10)$$

Invoking the normalized rms relation,  $D(0) \equiv R(1) = 1$ , produces  $(2\pi)^{-1} \int \phi(\beta) d\beta = 1$ , establishing that  $\phi(\beta)$  does indeed retain the same meaning in the scale-invariant formulation that it has in the classical formulation. The Mellin transform has already found many viable applications in the signal processing community (see, for example, Altes (1978), Casasent and Psaltis (1976 and 1977), Casasent and Kraus (1978)).

## 8. Concluding Remarks

The invariance property considered herein is important not only because it is characteristic of some as yet undetermined conservation principle, but also because it unifies processes which are traditionally stationary to those which are traditionally nonstationary. For example, the already formulated  $R(\alpha)$  unifies certain types of nonstationary random processes with classical stationary processes, since it is identical for both. Note that the ratio of  $t_2$  and  $t_1$  is the same whether  $L$  is a function of time or is in fact constant. In other words in  $R(\alpha)$  there is no distinction between a stationary random process and a self-similar or self-preserving nonstationary random process. This scale-independence reinforces the belief that since length scales are always defined by the analyst in some artificial or subjective manner, the actual physical phenomenon under consideration should nonetheless exhibit features which are natural to the phenomenon itself and independent of the scale chosen by the observer. Such behavior, by the way, is contrary to the notion of twentieth century physics that a particular numerical result

should always depend upon the relation of observed phenomena to observer. Invariance, then, is perhaps the major conceptual tool the use of which the scientist/engineer exploits the unity and interrelation of all stochastic behavior. In daily professional endeavors the analyst is typically unaware of this unity, and rather naturally tends to divide random processes into separate classes. This division is indeed useful and often necessary, but it is nonetheless an artifact of the "rational" type of knowledge, and not an intrinsic feature of nature (which more closely adheres to a principle of *plenitude*). Divisions, comparisons, categorizations, etc., create a world of distinctions, and any information about reality gleaned from such procedures is inherently limited. In conceptualizing about nature the scientist is faced with the same kind of problem as the three blind men who touched the elephant; only an approximate representation of reality can be expected.

More important, though, is the eduction of clear evidence that even though the statistical structure of a particular nonstationary random process may evolve in time in virtually any fashion, there exists nonetheless some well-defined invariance to the entire phenomena. This invariance is what stimulates the quest for those physical laws which bridge the gap between stationary and nonstationary behavior. Thus within the most disorderly of phenomena, there undoubtedly resides an underlying order co-existing with the disorder. In a universe governed by *entropy*, which stipulates that everything tends toward greater and ever greater disorder, and from which there appears to be no means of appeal, how does such order arise, why does it persist, and what are the as yet undiscovered laws which explain it? When formulated, such laws will answer these questions and simultaneously provide the "missing link" between these two species of phenomena.

## References

- Andreas, E. L. & Treviño, G., This volume.  
 Altes, R. A., *J. Acoust. Soc. Amer.* **63** (1978) 174.  
 Baudelaire, P., *Proc. IEEE* **60** (1974) 467.  
 Casasent, D. & Psaltis, D., *Appl. Opt.* **15** (1976) 1795.  
 Casasent, D. & Psaltis, D., *Proc. IEEE* **65** (1977) 77.  
 Casasent, D. & Kraus, M., *Appl. Opt.* **17** (1978) 1559.  
 Bendat, J. S. and Piersol, A. G., *Random Data: Analysis and Measurement Procedures* (Wiley-Interscience, New York, 1971).  
 Bhagavan, C. S. K., *Nonstationary Processes, Spectra and Some Ergodic Theorems* (Andhra University Press, Waltair, 1974).  
 Caloyannides, M. A., *J. Appl. Phys.* **45** (1974) 307.  
 Courant, R. and Hilbert, D., *Methods of Mathematical Physics I* (Interscience Publishers, New York, 1953).  
 Grace, J. H. and Young, A., *Algebra of Invariants* (Cambridge, 1903).  
 Gray, H. L. and Zhang, F. (1988). *J. Time Ser. Anal.* **9** (1988) 133.

- Gurevich, G. B., *Foundations of the Theory of Algebraic Invariants* (P. Noordhoff, Ltd. Groningen, The Netherlands, 1964).
- Hooge, F. N., *Physica B* **83** (1976) 14.
- Keshner, M. S., *Proc. IEEE* **70** (1982) 212.
- Keyser, C. J., *The Hibbert Journal* **3** (1904-05) 300.
- Lampard, D. G., *J. Appl. Phys.* **25** (1954) 802.
- Leighton, T.G., *Eur. J. Phys.* **9** (1988) 69.
- Levin, M.J., *IEEE Trans. Inf. Th.* **IT-10** (1964) 95..
- Loynes, R.M. (1968). *J. Roy. Statist. Soc. B* **30** (1968) 1.
- Nagabhushanam, K. and Bhagavan, C. S. K., *Can. J. Math.* **20** (1968) 1203.
- Page, C.H., *J. Appl. Phys.* **23** (1952) 103.
- Papoulis, A., *Probability, Random Variables, and Stochastic Processes* (McGraw-Hill, New York, 1965).
- Priestley, M.B., *J. Roy. Statist. Soc. B* **27** (1965) 204.
- Priestley, M.B., *J. Sound Vib.* **6** (1967) 86.
- Salmon, G., *Lessons Introductory to the Modern Higher Algebra* (Hodges, Figgis, and Co., Dublin, 1885).
- Turner, C. H. M., *J. Appl. Phys.* **25** (1954) 1347.
- Wornell, G. W., *IEEE Trans. Sig. Proc.* **40** (1992) 611.